

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

| | | |
|---|-----------|--|
| <p>(51) International Patent Classification ⁶ : C12N 15/52, 9/00, A61K 38/43, C07K 16/40, C12N 15/11, A61K 48/00, G01N 33/50</p> | <p>A1</p> | <p>(11) International Publication Number: WO 99/24583</p> |
| | | <p>(43) International Publication Date: 20 May 1999 (20.05.99)</p> |
| <p>(21) International Application Number: PCT/CA98/01059</p> <p>(22) International Filing Date: 12 November 1998 (12.11.98)</p> <p>(30) Priority Data: 2,220,805 12 November 1997 (12.11.97) CA 2,230,991 11 May 1998 (11.05.98) CA</p> <p>(71) Applicant (for all designated States except US): MCGILL UNIVERSITY [CA/CA]; 845 Sherbrooke Street West, Montréal, Québec H3A 2T5 (CA).</p> <p>(72) Inventors; and (75) Inventors/Applicants (for US only): SZYF, Moshe [CA/CA]; 7522 Wavell, Côte Saint-Luc, Québec H4W 2L7 (CA). BHATTACHARYA, Sanjoy [IN/CA]; Apartment 1403, 2150 Mackay Street, Montréal, Québec H3G 2M2 (CA). RAMCHANDANI, Shyam [CA/CA]; 5843 Hillcrest Crescent, Niagara Falls, Ontario L2J 2A8 (CA).</p> <p>(74) Agents: CÔTÉ, France et al.; Swabey Ogilvy Renault, Suite 1600, 1981 McGill College Avenue, Montréal, Québec H3A 2Y3 (CA).</p> | | <p>(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</p> <p>Published With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</p> |
| <p>(54) Title: DNA DEMETHYLASE, THERAPEUTIC AND DIAGNOSTIC USES THEREOF</p> | | |
| <p>(57) Abstract</p> <p>The present invention relates to a DNA demethylase enzyme having about 40 KDa, and wherein said DNA demethylase enzyme is overexpressed in cancer cells and not in normal cells. The present invention also relates to the therapeutic and diagnostic uses of the DNA demethylase.</p> | | |

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|----|--------------------------|----|--|----|--|----|--------------------------|
| AL | Albania | ES | Spain | LS | Lesotho | SI | Slovenia |
| AM | Armenia | FI | Finland | LT | Lithuania | SK | Slovakia |
| AT | Austria | FR | France | LU | Luxembourg | SN | Senegal |
| AU | Australia | GA | Gabon | LV | Latvia | SZ | Swaziland |
| AZ | Azerbaijan | GB | United Kingdom | MC | Monaco | TD | Chad |
| BA | Bosnia and Herzegovina | GE | Georgia | MD | Republic of Moldova | TG | Togo |
| BB | Barbados | GH | Ghana | MG | Madagascar | TJ | Tajikistan |
| BE | Belgium | GN | Guinea | MK | The former Yugoslav Republic of Macedonia | TM | Turkmenistan |
| BF | Burkina Faso | GR | Greece | ML | Mali | TR | Turkey |
| BG | Bulgaria | HU | Hungary | MN | Mongolia | TT | Trinidad and Tobago |
| BJ | Benin | IE | Ireland | MR | Mauritania | UA | Ukraine |
| BR | Brazil | IL | Israel | MW | Malawi | UG | Uganda |
| BY | Belarus | IS | Iceland | MX | Mexico | US | United States of America |
| CA | Canada | IT | Italy | NE | Niger | UZ | Uzbekistan |
| CF | Central African Republic | JP | Japan | NL | Netherlands | VN | Viet Nam |
| CG | Congo | KE | Kenya | NO | Norway | YU | Yugoslavia |
| CH | Switzerland | KG | Kyrgyzstan | NZ | New Zealand | ZW | Zimbabwe |
| CI | Côte d'Ivoire | KP | Democratic People's Republic of Korea | PL | Poland | | |
| CM | Cameroon | KR | Republic of Korea | PT | Portugal | | |
| CN | China | KZ | Kazakhstan | RO | Romania | | |
| CU | Cuba | LC | Saint Lucia | RU | Russian Federation | | |
| CZ | Czech Republic | LI | Liechtenstein | SD | Sudan | | |
| DE | Germany | LK | Sri Lanka | SE | Sweden | | |
| DK | Denmark | LR | Liberia | SG | Singapore | | |
| EE | Estonia | | | | | | |

DNA DEMETHYLASE, THERAPEUTIC AND
DIAGNOSTIC USES THEREOF

BACKGROUND OF THE INVENTION

5 (a) Field of the Invention

The invention relates to a novel enzyme, DNA demethylase, therapeutic and diagnostic uses thereof.

(b) Description of Prior Art

Many lines of evidence have established that
10 modification of cytosine moieties residing in the dinucleotide sequence CpG in vertebrate genomes is involved in regulating a number of genome functions such as parental imprinting, X-inactivation, suppression of methylation of ectopic genes and differential gene
15 expression (Szyf, M. (1996) Pharmacol. Ther. 70, 1-37). DNA methylation performs its function of differentially marking genes because the distribution of methylated CpGs is tissue- and site- specific forming a pattern of methylation (Szyf, M. (1996) Pharmacol. Ther. 70, 1-
20 37). It is clear that the pattern of methylation is fashioned by a sequence of methylation and demethylation events (Brandeis, M. et al. (1993) Bioassays 15, 709-713) during development and is maintained in the fully differentiated cell (Razin, A. et al. (1980) Science 210, 604-610). While it was originally suggested
25 that DNA demethylation is accomplished by a passive loss of methyl groups during replication (Razin, A. et al. (1980) Science 210, 604-610), it is now clear that an active process of demethylation occurs in embryonal
30 cells (Frank, D. et al. (1991) Nature 351, 239-241), in differentiating cell lines (Razin, A. et al. (1986) Proc. Natl. Acad. Sci. USA 83, 2827-2831; Szyf, M. et al. (1985) Proc. Natl. Acad. Sci. USA 82, 8090-8094) and in response to estrogen treatment (Saluz, H.P. et
35 al. (1986) Proc. Natl. Acad. Sci. USA 83, 7167-7171). Two modes of demethylation have been documented: site

specific demethylation that coincides in many instances with onset of gene expression of specific genes and a general genome wide demethylation that occurs during early development *in vivo* during cellular differentiation and in cancer cells (Feinberg, A.P. et al. (1983) Nature 301, 89-92; Razin, A. et al. (1986) Proc. Natl. Acad. Sci. USA 83, 2827-2831). The global demethylation is consistent with the hypothesis that a general demethylase activity which is activated at specific points in development or oncogenesis exists. It has been hypothesized that one mechanism regulating the pattern of methylation is the control of expression of methyltransferase (Szyf, M. (1991) Biochem. Cell Biol. 69, 764-767) and demethylase activities (Szyf, M. (1994) Trends Pharmacol. Sci. 7, 233-238). Although extensive information has been obtained on the enzymatic activity responsible for methylation and the regulation of its expression in the last two decades (Szyf, M. (1996) Pharmacol. Ther. 70, 1-37), the identity of the demethylase has remained a mystery. It is clear however that to fully understand how patterns of methylation are formed and maintained and to determine their role in development, physiology and oncogenesis, one has to identify the demethylase enzyme(s). Two main difficulties have inhibited the identification of this enzyme. First, it is believed that demethylation of a methylated cytosine is chemically highly unlikely since it involves breaking a very stable C-C bond. Second, demethylation occurs at very defined stages in development (Brandeis, M. et al. (1993) Bioassays 15, 709-713) and identifying an adequate tissue source for this enzyme is critical.

Whereas no *bona fide* demethylase has been identified to date, alternative biochemical mechanisms involving exchange of methylated cytosines with non-

5 methylated cytosines have been described. One previously proposed mechanism is removal of the methylated base by a glycosylase and its replacement with a non-methylated nucleotide utilizing an "excision-repair" mechanism (Razin, A. et al. (1986) Proc. Natl. Acad. Sci. USA 83, 2827-2831). Glycosylase activities that can remove methylated cytosines from DNA have been demonstrated by Vairapandi and Duker (Vairapandi, M. et al. (1993) Nucl. Acids Res. 21, 5323-5327) and more recently by Jost (Jost, J. P. et al. (1995) J. Biol. Chem. 270, 9734-9739). However it is not clear whether this activity is responsible for the general demethylation observed in cellular differentiation. The fact that the activity identified by Jost acts specifically on hemimethylated sequences (which is not the natural substrate in most cases) and can remove thymidines as well as 5-methylcytosines, supports a repair function for this glycosylase-demethylase (Jost, J. P. et al. (1995) J. Biol. Chem. 270, 9734-9739). An alternative mechanism involving a RNA dependent activity has been recently described by Weiss et al. (Weiss et al., 1996). This proteinase-insensitive RNA dependent activity has been shown to catalyze the excision and replacement of a methylated CpG dinucleotide with a nonmethylated CpG dinucleotide that is contained in a DNA-RNA hybrid molecule (Weiss, A. et al. (1996) Cell 87, 709-718). This activity which was identified in differentiating cells in culture was proposed to be involved in demethylation during development. These previous findings demonstrate that the common accepted model in the field has been that a *bona fide* demethylase does not exist.

It has been previously proposed that the extensive hypomethylation observed in cancer cells might be a consequence of activation of demethylase activity by

oncogenic pathways (Szyf, M.(1994) Trends Pharmacol. Sci. 7, 233-238; Szyf, M. et al. (1995) J. Biol. Chem. 270, 12690-12696). In accordance with this hypothesis we have shown that ectopic expression of v-Ha-ras had
5 induced demethylation activity in the cells (Szyf, M. et al. (1995) J. Biol. Chem. 270, 12690-12696). Using an assay that directly measures the conversion of 3'³²P labeled methyl dCMP (mdCMP) into dCMP, we have shown that nuclear extracts prepared from P19-Ras transfectants bear high levels of demethylase activity (Szyf,
10 M. et al. (1995) J. Biol. Chem. 270, 12690-12696). Building on this observation, we hypothesized that cancer cell lines were a good source for demethylase. However, it is not evident that Ras expression in p19
15 cells does reflect the situation in cancer cells. P19 is an embryonic cell and expression of Ras might be differentiating them.

It would be highly desirable to be provided with a *bona fide* DNA demethylase (DNA dMTase) to alter
20 developmental programs for therapeutic and biological use.

SUMMARY OF THE INVENTION

In accordance with the present invention, we
25 demonstrate the purification of a *bona fide* DNA demethylase (DNA dMTase) from a human lung cancer cell line A549, determine its kinetic parameters and substrate specificity. The DNA dMTase activity identified in this study converts methyl-dCMP (mdCMP) residing in the
30 dinucleotide sequence mdCpG into dCMP whereas the methyl group is released as a volatile residue which was identified to be methanol. The activity is purified away from any trace amounts of dCTP, is insensitive to the DNA polymerase inhibitor ddCTP, is not
35 affected by the presence of methyl dCTP (mdCTP) in the

reaction and does not exhibit exonuclease or glycosylase activities. The identification of this new enzyme points out to new directions in our understanding of how DNA methylation patterns are formed and
5 altered.

One aim of the present invention is to provide a *bona fide* DNA demethylase (DNA dMTase).

In accordance with the present invention there is provided a DNA demethylase enzyme having about
10 40 KDa, and wherein the DNA demethylase enzyme is over-expressed in cancer cells and not in normal cells.

In accordance with the present invention there is provided a cDNA encoding human demethylase which comprises a sequence set forth in SEQ ID NO:1.

15 In accordance with the present invention there is provided two mouse cDNAs homologous to the human cDNA, wherein the cDNA encoding mouse demethylase having a sequence set forth in SEQ ID NOS:5-7.

In accordance with the present invention there
20 is provided a different human cDNA which encodes a protein homologous to the human demethylase having a sequence set forth in SEQ ID NO:3.

In accordance with the present invention there is provided the use of the expression of demethylase
25 cDNAs to alter DNA methylation patterns of DNA *in vitro* in cells or *in vivo* in humans, animals and in plants.

The demethylase cDNAs expression may be under the direction of mammalian promoters, such as CMV.

The demethylase cDNAs expression may be under
30 plant specific promoters to alter methylation in plants and to allow for altering states of development of plants and expression of foreign genes in plants.

The demethylase cDNAs expression may be in the antisense orientation to inhibit demethylase in cancer
35 cells for therapeutic processes.

The expression of demethylase cDNA in mammalian cells may be to alter their differentiation state and to generate stem cells for therapeutics, cells for animal cloning and to improve expression of foreign genes.

5 In accordance with the present invention there is provided the use of the expression of demethylase cDNAs in bacterial or insect cells for production of large amounts of demethylase.

10 In accordance with the present invention there is provided the use of the expression of demethylase cDNAs for the production of protein in vertebrate, insect or bacterial or plant cells, such as antibodies against demethylase.

15 In accordance with the present invention there is provided the use of the sequence of demethylase cDNAs as a template to design antisense oligonucleotides and ribozymes.

20 In accordance with the present invention there is provided the use of the predicted peptide sequence of demethylase cDNAs to produce polyclonal or monoclonal antibodies against demethylase.

25 In accordance with the present invention there is provided the use of expression of cDNAs in two hybrid systems in yeast to identify proteins interacting with demethylase for diagnostic and therapeutic purposes.

30 In accordance with the present invention there is provided the use of expression of cDNAs in bacterial, vertebrate or insect cells to produce large amounts of demethylase for obtaining a x-ray crystal structure and for high throughput screening of demethylase inhibitors for therapeutics and biotechnology.

In accordance with the present invention there is provided a volatile assay for high throughput

screening of demethylase inhibitors as therapeutics and anticancer agents which comprises the steps of:

- a) using transcribed and translated demethylase cDNAs *in vitro* to convert methyl-cytosine present in methylated DNA samples to cytosine present in DNA and volatilize methyl group;
- b) determining the absence or minute amount of volatilize methyl group as an indication of an active demethylase inhibitor.

In accordance with the present invention there is provided a volatile assay for the diagnostics of cancer in a patient sample which comprises the steps of:

- a) determining demethylase activity in patient samples by assaying conversion of methyl-cytosine present in methylated DNA to cytosine present in DNA and its volatilization as methyl groups released as methanol;
- b) determining the presence or minute amount of volatilized methyl released as methanol groups as an indication of cancer in the patient sample.

In accordance with the present invention there is provided the use of an antagonist or inhibitor of DNA demethylase for the manufacture of a medicament for cancer treatment, for restoring an aberrant methylation pattern in a patient DNA, or for changing a methylation pattern in a patient DNA.

Such an antagonist is a double stranded oligonucleotide that inhibits demethylase at a K_i of 50nM, such as $[C^mGC^mGC^mGC^mG] \cdot [G^mCG^mCG^mCG^mC]_n$.

The inhibitors include, without limitation an anti-DNA demethylase antibody, an antisense of DNA demethylase or a small molecule such as any derivative of imidazole.

The change of the methylation pattern may activate a silent gene. Such an activation of a silent gene permits the correction of genetic defect such as found for β -thalassemia or sickle cell anemia.

5 The DNA demethylase of the present invention may be used to remove methyl groups on DNA *in vitro* such as needed for cloning DNA.

10 The DNA demethylase of the present invention or its cDNAs may be used, for changing the state of differentiation of a cell to allow gene therapy, stem cell selection or cell cloning.

The DNA demethylase of the present invention or its cDNAs may be used, for inhibiting methylation in cancer cells using vector mediated gene therapy.

15 In accordance with the present invention there is provided an assay for the diagnostic of cancer in a patient, which comprises determining the level of expression of DNA demethylase by either RT-PCT, ELISA or volatilization assay of the present invention in a
20 sample from the patient, wherein overexpression of the DNA demethylase is indicative of cancer cells.

BRIEF DESCRIPTION OF THE DRAWINGS

25 Figs. 1A to 1B illustrate the purification of demethylase (DNA dMTase) from human A549 cells;

Figs. 2A and 2C illustrate that DNA dMTase is a protein inhibited by RNA and not by ddCTP, mdCTP;

Figs. 2B and 2D illustrate the kinetics of DNA dMTase activity;

30 Figs. 3A to 3C illustrate the product of DNA dMTase activity is cytosine and it exhibits no exonuclease or glycosylase activity;

Figs. 4A-4C illustrate the demethylation reaction releases methanol as a volatile residue;

Fig. 4D illustrates the transfer of a proton from water to regenerate cytosine;

Figs. 4E-4F illustrate that the volatile product is methanol;

5 Fig. 5 illustrates the suggested demethylation reaction;

Figs. 6A-6D illustrate the substrate Specificity of DNA dMTase;

10 Figs. 7A-7D illustrate chromatographic isolation of dMTase from human A549 cells;

Figs. 8A-8B illustrate the alignment between the MDB domain of MeCP2 and demethylase and the predicted amino acid sequence of human demethylase;

15 Fig. 8C illustrates the mRNA encoded by demethylase;

Figs. 9A-9F illustrate the cDNA and their predicted amino acid of demethylases and homologues of the present invention (SEQ ID NOS:1-8);

20 Figs. 10A-B illustrate a mammalian expression vector of dMTase and *in vitro* translated dMTase polypeptide;

Fig. 10C illustrates that *in vitro* translated DNA dMTase releases volatile methyl residues from methylated DNA;

25 Fig. 10D illustrates that *in vitro* translated DNA dMTase transform methylated cytosines to cytosines;

Fig. 11A illustrates that transiently transfected demethylase releases volatile residues from methylated DNA;

30 Fig. 11B illustrates the polypeptide expressed from transiently transfected demethylase;

Figs. 11C-11E illustrate that transiently transfected demethylase transforms methylated cytosines to cytosines in a protein dependent manner;

Fig. 11F illustrates that the transformation of methylated cytosine to cytosine by transiently transfected demethylase depends on the concentration of substrate;

5 Fig. 12A illustrates that transiently transfected demethylase catalyzes the transfer of a proton from tritiated water to regenerate cytosine;

Fig. 12B illustrates that the cloned demethylase releases methanol from methylated DNA;

10 Figs. 13A-13C illustrate that the cancer cells express demethylase activity whereas normal cells do not;

Fig. 13D illustrates that demethylase mRNA is highly express in cancer cells;

15 Fig. 14A illustrates demethylase bacterial retroviral and mammalian expression vector;

Fig. 14B illustrates inhibition of demethylase activity by a specific inhibitor;

20 Fig. 14C illustrates inhibition of tumorigenesis *in vitro* by an inhibition of demethylase;

Fig. 15 illustrates inhibition of tumorigenesis in cell culture by induced expression of demethylase antisense vector;

25 Fig. 16 illustrates the inhibition of demethylase by a small molecule inhibitor imidazole; and

Fig. 17 illustrates a model for the inhibition of cancer growth by an inhibition of demethylase.

DETAILED DESCRIPTION OF THE INVENTION

30 The pattern of methylation is fashioned during development by a sequence of methylation and demethylation events. The identity of the demethylase has remained a mystery and alternative biochemical activities have been shown to demethylate DNA but no activity
35 that can truly remove methyl groups from DNA has been

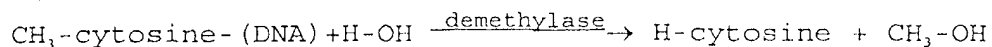
shown to date. Utilizing human lung carcinoma cells as a source for demethylase activity we demonstrate that mammalian cells bear a *bona fide* DNA demethylase (DNA dMTase) activity. DNA dMTase transforms methyl-C to C by catalyzing replacement of the methyl group on the 5 position of C with a hydrogen derived from water. DNA dMTase demethylates both fully methylated and hemimethylated DNA, shows dinucleotide specificity and can demethylate mdCpdG sites in different sequence contexts. This enzyme is different from previously described demethylation activities: it is proteinase sensitive, activated by RNase and releases different products.

DNA dMTase is a novel enzyme showing a new and unexpected activity that has not been previously described in any organism. The finding of a *bona fide* demethylase, points out new directions in our understanding of the biological role of DNA methylation.

In spite of the fact that it was previously shown that Ras expression in p19 cells can induce demethylation activity. It was not clear whether this demethylation activity is indeed a *bona fide* demethylase. One would predict that demethylase is present in embryonal cells. It was surprising to see that demethylation activity is present in cancer cells. The finding of high levels of demethylase in A549 cells is indeed an unexpected discovery.

In accordance with the present invention, it is shown and demonstrated that demethylation occurs by removal of a methyl group from methylated cytosine in DNA, that a hydrogen from water replaces the methyl group at the 5' position, that the resulting methyl group reacts with the remaining hydroxyl from water to generate methanol which volatilizes (Fig. 4E-F). Thus,

bona fide demethylation of DNA involves the following reaction:



5 The cDNA cloned in accordance with the present invention is the demethylase since it can convert methyl-cytosines in DNA to cytosines and volatilize the methyl groups on DNA when transcribed and translated *in vitro* which are released as methanol. This is a novel
10 cDNA encoding a biochemical activity that has been not described before.

In accordance with the present invention, there is shown a model for the inhibition of cancer growth by an inhibition of demethylase (Fig. 17).

15

EXPERIMENTAL PROCEDURES

Cell Culture

A549 Lung Carcinoma cells (ATCC: CCL 185) were grown in Dulbecco's modified Eagle's medium (with low
20 glucose) supplemented with 10% fetal calf serum, 2 mM glutamine, 10 U/ml ciprofloxacin. Human Skin Fibroblasts #72-213A MRHF were obtained from BioWhittaker, Bethesda and were grown in Dulbecco's modified Eagle's medium supplement with 2% fetal calf serum, 2 mM glutamine.
25 H446 Lung carcinoma cells (ATCC: HTB 171) was grown in RPMI 1640 medium with 5% fetal calf serum.

Preparation of nuclear extract

Nuclear extracts were prepared from A549 cultures at near confluence as previously described (Szyf
30 et al., 1991; Szyf et al., 1995). The cells were trypsinized, collected and washed with phosphate-buffered saline and suspended in buffer A (10 mM Tris, pH 8.0, 1.5 mM MgCl₂, 5mM KCl, 0.5% NP-40) at the concentration of 10⁸ cells per ml for 10 min. at 4°C. Nuclei were
35 collected by centrifugation of the suspension at 1000 g

for 10 minutes. The nuclear pellet was resuspended in buffer A (400 μ l) and collected as described in the experimental procedures. A nuclear extract was prepared from the pelleted nuclei by suspending them in
5 buffer B (20 mM Tris, pH 8.0, 25% glycerol, 0.2 mM EDTA and 0.4 mM NaCl) at the concentration of 3.3×10^8 nuclei per ml and incubating the suspension for 15 min. at 4°C. The nuclear extract was separated from the nuclear pellet by centrifugation at 10,000g for 30 min-
10 utes. Nuclear extract were stored in -80°C for at least two months without loss of activity.

Chromatography on DEAE-Sephadex

A freshly prepared nuclear extract (1 ml , 1.1 mg) was passed through a Microcon™ 100 spin column, the
15 retainant was diluted to a conductivity equivalent to 0.2 M NaCl in buffer L and applied onto a DEAE-Sephadex column (Pharmacia) (1.0 x 5 cm) that was preequilibrated with buffer L (10 mM Tris-HCl, pH 7.5, 10 mM MgCl₂) containing 0.2 M NaCl at a flow rate of 1
20 ml/min. The column was then washed with 15 ml of the starting buffer (buffer L + 0.2 M NaCl) and proteins were eluted with 5 ml of a linear gradient of NaCl (0.2-5.0 M). 0.8 ml fractions were collected and assayed for demethylase activity after desalting
25 through a Microcon™ 10 spin column (Amicon) and resuspension of the retainant in 0.8 ml buffer L. DNA demethylase eluted between 2-5.0 M NaCl.

Chromatography on S-Sepharose

Active DEAE-Sepharose column fractions were
30 pooled, adjusted to 0.1 M NaCl by dilution and loaded onto an S-sepharose column (Pharmacia) (1.0 x5 cm) which had been preequilibrated with buffer L containing 0.2 M NaCl at a flow rate of 1 ml/min. Following washing of the column as described in experimental proce-
35 dures, the proteins were eluted with 5 ml of a linear

NaCl gradient (0.2-5.0M). 0.5 ml fractions were collected and assayed for DNA demethylase activity after desalting and concentrating to 0.2 ml using a Microcon™ 10 spin column. DNA demethylase activity eluted around 5.0 M NaCl.

Chromatography on Q-Sepharose

Active fractions from S-sepharose column were pooled, adjusted to 0.2 M NaCl by dilution and applied onto a Q-sepharose (Pharmacia) column (1.0 x5 cm) which had been equilibrated as described in the experimental procedures at a flow rate of 1 ml/min. The column was washed and the proteins were eluted with a linear NaCl gradient (0.2- 5.0 M). Fractions (0.5 ml) were collected, assayed for demethylase activity after desalting and concentrating to a final volume of 0.2 ml as described in the experimental procedures. The demethylase activity eluted around 4.8-5.0 M NaCl.

Gel-Exclusion Chromatography on DEAE-Sephacel

The pooled fractions of Q-sepharose column were adjusted to 0.2 M NaCl, loaded onto a 2.0 x 2.0 cm DEAE-Sephacel column (Pharmacia) and eluted with 10 ml of buffer L containing 0.2 M NaCl. The fractions (0.8 ml) were collected and assayed after concentration to about 180 µl with a Microcon™ 10 spin column for DNA demethylase activity. The activity was detected at fraction 4, which is very near the void volume (~200kDa).

Assay of DNA demethylase activity

To directly assay DNA demethylase activity in vitro two independent methods were applied.

(A) To assay the conversion of methyl-dCMP (mdCMP) to dCMP we used a previously described method (Szyf et al., 1995). Briefly, $\alpha^{32}\text{P}$ labeled, fully methylated poly[mdC³²PdG]n substrate was prepared as follows. One hundred ng of a double-stranded fully methylated

(mdCpdG) oligomer (Pharmacia) were denatured by boiling, which was followed by partial annealing at room temperature. The complementary strand was extended with Klenow fragment (Boehringer Mannheim) using methyl-5-dCTP (mdCTP, 0.1 mM) (Boehringer Mannheim) and $[\alpha\text{-}^{32}\text{P}]$ GTP (100 μCi , 3000 Ci/mmol), and the unincorporated nucleotides were removed by chromatography through a NAP-5 column (Pharmacia). The NAP-5 chromatography was repeated to exclude minor contamination with unincorporated nucleotides. As a control a non-methylated poly[dC³²pdG]n substrate was similarly prepared except that a nonmethylated dCpdG oligomer served as a template and dCTP was used in the extension reaction. The column fractions (30 μl), described in the experimental procedures were incubated with 1 ng of poly[mdC³²pdG]n substrate for 1 hour at 37°C in a buffer L containing 25% glycerol (v/v) and 5 mM EDTA. The reacted DNA as well as a nonmethylated poly[dC³²pdG]n and methylated [mdC³²pdG]n nonreacted controls were purified by phenol/chloroform extraction and subjected to micrococcal nuclease digestion (100 μg at 10 μl) and calf spleen phosphodiesterase (2 μg) (Boehringer) (Pharmacia) to 3' mononucleotides for 15 hours at 37°C. The digestion products were loaded onto a thin layer chromatography plate (TLC) (Kodak, 13255 Cellulose), separated in a medium containing, 132 ml Isobutyric acid:40 ml water: 4 ml ammonia solution, autoradiographed and the intensity of the different spots was determined using a phosphorimager (Fuji, BAS 2000). ³²P labeled substrates and tritium labeled substrates were phosphoimaged using BAS 2000 plate and BAS-TR2040 phosphorimager plate respectively.

(B) The second method determined removal of methylated residues from methylated DNA by measuring disappearance of ³H-CH₃ or ¹⁴C-CH₃ from the reaction mixture. 100 ng

of poly [dCdG]n double stranded DNA was methylated using SssI methylase (New England Biolabs) and an excess of [³H-methyl AdoMet (80 Ci/mmol; New England Nuclear)]. The tritiated methyl group containing DNA was purified from labeled AdoMet using NAP-5 column chromatography. All column purified fractions of DNA demethylase were assayed using the tritiated substrate. In a typical assay, 1 ng of DNA was incubated (at a specific activity of 4 x10⁶dpm/mg) with 30 µl of column fraction for one hour at 37°C in buffer L. To determine the number of methyl groups remaining in the DNA following incubation with the different fractions, 250 µl of water were added and the mixture was incubated at 65°C for 5 minutes. One hundred µl of the reaction mixture were withdrawn for liquid scintillation counting. Controls received similar treatment except that in place of a column fraction, an equal volume of buffer L was added. The number of methyl groups that were removed from the DNA by the different fractions was determined by subtracting the remaining counts in each of the fractions from the counts remaining in the control. All tests were carried out in triplicates. The results are presented as picomole methyl group removed. One unit of DNA dMTase activity is defined as: amount of enzyme that releases one picomole of methyl group from methylated dCpdG substrate in one hour at 37°C.

Methyl removal assay using double-labeled substrates

To determine whether the methyl group leaves the DNA and not any non-specific removal of tritium, we prepared SK plasmid DNA containing a tritiated hydrogen at the 6' position of cytosine and thymidine by growing the plasmid harboring bacteria in the presence of deoxy [6-³H] Uridine (22 Ci/mmol; Amersham) (10µCi/ml). The [6-³H]-cytosine containing pBluescript SK(+) was puri-

fied according to standard protocols and was methylated using an excess of [^{14}C -methyl] AdoMet (59 mCi/mmol; Amersham) (10 μCi per 100 μl reaction) and SssI methylase. The double labeled DNA substrate was purified
5 twice on a NAP-5 column. 15 μl of DNA dMTase were incubated with 1 ng of double labeled DNA (specific activity of 2000 dpm/ng) for 1 hour at 37°C. Following incubation, the remaining ^{14}C versus ^3H counts were determined as described in the experimental procedures
10 by scintillation counting (Wallac). The ^{14}C counts were normalized against ^3H counts. The controls received similar treatment except that instead of DNA dMTase, an equal amount of distilled water was added to them.

To determine the number of ^3H -CH₃ in the gaseous
15 phase, 1 ng of ^3H -CH₃ poly [dCpdG] DNA were incubated with DNA dMTase overnight in a sealed tube (Pierce, Illinois, USA). 0.8 ml of air were removed from the tube using a gas tight syringe (Hamilton, Reno, Nevada) and injected into a sealed gas tight scintillation vial
20 containing 10 ml OptiPhase scintillation fluid (Wallac, UK) and counted. As a control the DNA was incubated with an equal volume of buffer L and treated similarly.

Synthesis of other methylated dC dinucleotides

Poly [mdC³²pdA] and [mdC³²pdT] substrates were
25 prepared as follows. About 0.5 μg of 20 mer oligonucleotides 5'(GG)103', 5'(GT)103' and 5'(GA)103' were boiled and annealed at room temperature with oligonucleotide 5'CCCCCC3', 5'CACACA3' and 5'CTCTCT3' respectively. The complementary strand was extended with
30 Klenow fragment using m5dCTP (Boehringer Mannheim) and either [$\alpha^{32}\text{P}$] dATP (100 μCi , 3000 Ci/mmol) or [$\alpha^{32}\text{P}$] dTTP (100 μCi , 3000 Ci/mmol) respectively. The unincorporated nucleotides were removed by chromatography through a NAP-5 column. Hemimethylated mdCpG substrate
35 was prepared in a similar manner except that a nonmeth-

ylated poly dCpdG substrate (Boehringer) was used as template and m5dCTP and [$\alpha^{32}\text{P}$]dGTP were used for extension as described in the experimental procedures.

Assay for nuclease and glycosylase activity

5 [$^{32}\text{pmdCpdG}$]n substrate which included a labeled ^{32}P 5' to mdC was prepared as follows. About 100 ng of poly dCpdG DNA were boiled and partially annealed at room temperature. [$\alpha^{32}\text{P}$]dCTP and cold dGTP were used for complementary strand extension as described in the
10 experimental procedures. The free nucleotides were separated using NAP-5 column chromatography. The purified [$^{32}\text{pmdCpdG}$]n DNA was subjected to methylation by SssI methylase using 320 μM AdoMet. The DNA was repurified twice using a NAP-5 column. The methylated DNA (1
15 ng) was incubated with either 30 μl DNA dMTase, nuclear extract or buffer L. To determine whether $\alpha^{32}\text{P}$ labeled residue is excised from the DNA it was directly applied (3 μl) onto a TLC plate. To determine whether the DNA was demethylated it was subjected to digestion with
20 snake venom phosphodiesterase (0.2 mg in a 10 μl reaction volume) (Boehringer Mannheim) which attacks the 3'-OH group releasing 5'-mononucleotides. The resulting mononucleotides were separated on TLC plates and autoradiographed.

25 To test whether dCTP copurifies with DNA dMTase, which may be involved in activities other than *bona fide* demethylation, 20 μM of dCTP with 1 μl of $\alpha^{32}\text{P}$ labeled dCTP (3000 Ci/mmole) was loaded onto the column with nuclear extract. The ^{32}P counts were measured in
30 the flow through, washes and in the different fractions. About 1.1 million counts were loaded onto the DEAE-Sepharose column and were all recovered up to fraction 8.

35 To determine whether DNA dMTase contains a DNA polymerase activity, DNA demethylase reactions were

performed in presence of 500 μ M of ddCTP (Pharmacia) or 500 μ M of m5dCTP (Boehringer Mannheim) at initial rate conditions.

To determine whether DNA dMTase is sensitive to
5 RNase or Proteinase K treatment, DNA dMTase was pre-treated for 1 h at 56°C with 200 μ g/ml proteinase K (Sigma). A demethylation reaction was carried out with this pretreated fraction in the usual manner using both
10 demethylation assays described in the experimental procedures. To test the effect of RNA digestion on the demethylation reaction, the fractions from different columns were treated with 100 μ g/ml RNase A (Sigma).

Demethylation of pBluescript SK(+) Plasmid

About 4 μ g plasmid pBluescript SK (Stratagene)
15 was subjected to methylation using SssI methylase. The methylated plasmid (4 ng) was incubated with 30 μ l of DNA dMTase Fraction 4 of DEAE-Sephacel column under standard conditions, extracted with phenol: chloroform and precipitated with ethanol. About 1 ng of the plas-
20 mid were subjected to digestion with 10 units each of either of the restriction endonucleases EcoRII (GIBCO-BRL), DpnI, HhaI or HpaII (New England Biolabs) before and after methylation as well as after DNA dMTase treatment in a reaction volume of 10 μ l for 2 hour at
25 37°C. Following restriction digestion the plasmids were extracted with phenol:chloroform, ethanol precipitated and resuspended in 10 μ l. The plasmids were electrophoresed on a 0.8% (w/w) Agarose gel, transferred onto a Hybond Nylon membrane and hybridized with
30 pBluescript SK(+) plasmid which was ³²P labeled by random-priming (Boehringer Mannheim).

Effect of Redox Reagents (NAD, NADH, NADP, NADPH and FeCl₃) on demethylase activity

The reagents were prepared at 100 μ M concentra-
35 tion and added at a final concentration of 10 μ M to a standard methyl removal assay under initial rate condi-

tions as described in the experimental procedures. The methyl removal activity in presence of each of the cofactors was compared to a control DNA dMTase reaction.

5 **Determination of kinetic parameters**

For determination of kinetic parameters, the demethylation reactions were performed using both assays (generation of dCMP and removal of methyl) as described in the experimental procedures except that
10 varying DNA concentrations from 0.1 nM to 2.5 nM were used in a total volume of 50 μ l including 30 μ l of DNA dMTase. Since it has been established by previous experiments that the reaction proceeds for at least 3 hours, the initial velocity of reaction was measured
15 at one hour intervals. The velocity data was collected at each substrate DNA concentration range stated for both assays. The K_m and V_{max} values for DNA demethylase activity were determined from double reciprocal plots of velocity versus substrate concentration.

20 **Measurements of methanol production catalyzed by demethylase by gas chromatography**

Gas chromatography was performed with a VarianTM model 3400 GC equipped with a 30m StabilwaxTM column (0.053 cm i.d.: Restek Corporation). NitrogenTM was
25 used as carrier gas at a flow rate of 32 ml/min, the injector and detector chambers were at 200 and 300°C respectively. The column was maintained at 40°C for 5 minutes after sample injection.

The demethylase reaction was performed in eppendorf tubes kept within sealed scintillation vials with
30 300 μ l of water as aqueous phase (in radioactive trapping experiments this was replaced by 300 μ l of methanol). The demethylase reaction was initiated in buffer L (10 mM $MgCl_2$, 10 mM Tris-HCl pH 8.0) with 500 ng of
35 tritiated SK plasmid (6000 dpm/ μ l) and 100 μ l of demethylase at 37°C. After overnight incubation at 37°C,

the aqueous phase surrounding the eppendorf tube was transferred to a fresh eppendorf tube, 2 μ l of this mixture was injected in the gas chromatography using a gas tight syringe (Hamilton, Reno, Nevada).

5 **Coupled *in vitro* transcription translation**

The mRNAs encoded by the pcDNA 3.1/His Xpress demethylase constructs described above were transcribed and translated by coupled transcription-translation using Promega™ TNT reticulocyte lysate kit (according
10 to manufacturer's protocol), 2 μ g of each construct and 40 μ Ci of [³⁵-S]methionine (1,000Ci/mmol, Amersham) in a 50 μ l reaction volume. To purify non labeled *in vitro* translated demethylase, coupled *in vitro* transcription and translation was performed as above but in the pres-
15 ence of cold methionine. The translation products were bound to a Probond™ nickel column (Invitrogen) and demethylase was eluted according to the manufacturer's protocol with increasing concentrations of imidazole. Demethylase is eluted at 350-500mM imidazole. The imi-
20 dazole eluted demethylase was dialyzed and concentrated by lyophilization.

25 **Gas chromatography coupled with Mass spectrometry (GC-MS) Analyses for identification of volatile product of demethylase catalyzed reaction as methanol**

The demethylation reactions (volume 50 μ l) were run in conical vials having a total internal volume of 350 microlitres. The vials were closed with a teflon-lined screw cap and left at room temperature for 18 h.
30 The vials were cooled in an ice bath, opened and 10 mg of NaCl and 50 microlitres of toluene were added. The vials were frequently shaken over a period of 1 h. The toluene phases were pipetted into clean vials in a manner to rigorously exclude water carry over. Anhydrous
35 sodium sulfate (5 mg) was added to the toluene extracts to remove water, and the toluene phases were pipetted

into autoinjector vials for GC/MS analysis. Aliquots of 3 microlitres were analyzed under the following instrumental conditions: Instrument: Hewlett-Packard 5988A; Column: 30 m x 0.25 mm i.d. fused quartz capillary with 0.25 micron DB-1 liquid phase, programmed after an initial hold for 1 min at 70 deg at 5 deg/min to 80 deg, then ramped ballistically to 280 deg for bake-out for 5 min; Injector and interface temperatures: 250 deg; Helium flow rate 1.5 ml/min; Mass spectrometer: ion source 200 deg, 70 eV electron impact ionization, scanning from m/z 10 to 50 in full scan mode was begun 6 s after injection, and ceased at 1.5 min to avoid acquisition of the intense toluene solvent peak.

Human A549 cells bear a demethylase activity that could be purified away from dCTP and DNA MeTase

The use of an appropriate cellular source and a direct assay for demethylase activity are obviously critical. As we have previously shown that demethylase activity was induced in response to ectopic expression of the Ras oncogene (Szyf et al., 1995) we reasoned that cancer cells might bear high levels of demethylase activity. Based on preliminary studies demonstrating the presence of high levels of demethylase activity in the human lung carcinoma cell line A549, we have chosen this cell line for our further studies and purification steps. Previous studies have used indirect measures such as increased sensitivity to methylation-sensitive restriction enzymes as indicators of demethylase activity (Weiss et al., 1996; Jost et al., 1995). To directly measure the conversion of 5-mdCMP in DNA to dCMP, we have utilized a completely methylated ³²P labeled [mdC³²pdG]n double stranded oligomer which we had previously described (Szyf et al., 1995). Following incubation with the different fractions, the

DNA is purified and subjected to cleavage with micrococcal nuclease to 3' mononucleotides. The 3' labeled mdCMP and dCMP are separated by thin layer chromatography (TLC) and the conversion of mdCMP to dCMP is directly determined. This assay provides a stringent test for *bona fide* demethylation and discriminates it from previously described 5mCpC replacement activities (Jost et al., 1995; Weiss et al., 1996). The glycosylase-demethylase activity described by Jost et al. (Jost et al., 1995) will require the presence of a ligase activity and an energy source for replacement of mdC with C to be detected by our assay, whereas the demethylase activity described by Weiss et al. will not be detected since it replaces the intact mdC³²pdG dinucleotide with a cold dCpdG without altering its state of methylation (Weiss et al., 1996).

Nuclear extracts were prepared from A549 cells, applied onto a DEAE-Sephadex column, eluted with a linear gradient from 0.2-5.0M NaCl and the fractions were assayed for demethylase (dMTase) activity as described in the experimental procedures. As shown in Fig. 1(A) a clear peak of dMTase activity is eluted at the high salt fraction 10.

Conversion of methylated cytosine to cytosine:
Nuclear extracts prepared from A549 cells (1.1 mg) were passed through an AMICON™ 100 spin column. The retainant (98.56 mg, 0.2 mg/ml) was loaded onto a DEAE-Sepharose column, the different chromatographic column fractions eluted by a linear NaCl gradient (0.2-5M) were desalted and (30 µl) incubated with 1 ng of [mdC³²pdG]n double stranded oligomer for 1 hour at 37°C, digested to 3' mononucleotides and analyzed on TLC as described in the experimental procedures. Control methylated (ME) and nonmethylated (NM) [dC³²pdG]n substrates were digested to 3' mononucleotides and loaded on the TLC

plate to indicate the expected position of dCMP and mdCMP. The active fraction is indicated by an arrow. This fraction was loaded on S-Sepharose followed by Q-Sepharose and DEAE-Sepharcel fractionation.

5 The first chromatography step purified the dMTase activity from the bulk of nuclear protein (Fig. 1B) and is a very effective purification step.

DNA dMTase activity as measured by the release of volatile methyl residues. The different column
10 fractions were incubated with 1ng (4×10^6 dpm/ μ g) of [3 H]-CH₃-[mdCpdG]n oligomer and the release of volatile methyl residues was determined (-) and presented as total dpm). The results are an average of three independent determinations. Protein concentration was
15 determined using the Bio-Rad Bradford kit (-). The elution profile of 20 μ M of [32 P]- α -dCTP incubated with the protein was determined by scintillation counting of the different DEAE fractions (-) and presented as fraction of dCTP loaded on the column.

20 To exclude the possibility that the DNA dMTase activity detected in our assay is carried by the DNA MeTase, we assayed the fractions for DNA MeTase activity using a hemimethylated DNA substrate as previously described (Szyf et al., 1991). As observed in Figure
25 1B DNA MeTase activity is detected in the second and third fractions, thus our fractionation separated DNA dMTase away from the DNA MeTase suggesting that they are independent proteins.

There is a remote possibility that the demethylation observed is not a *bona fide* demethylation but
30 a consequence of a glycosylase removal of mC, followed by removal of the remaining deoxyribose-phosphate by AP (apyrimidine) nuclease, repair of the gap catalyzed by DNA polymerase using trace dCTP contained in the fraction and ligation of the break with ligase in the pres-
35

ence of residual ATP. For this hypothesis to be consistent with our data, four independent enzymes and two cofactors have to cofractionate with DNA dMTase. To exclude the possibility that a trace amount of dCTP is bound to DNA dMTase active fraction, we have added 20 μ M of 32 P labeled dCTP (10×10^6 cpm) to the nuclear extract and determined its elution profile on the DEAE column. Less than background cpm (10 cpm) were detected in the DNA dMTase active fraction suggesting that our first column purifies dCTP away from the DNA dMTase at least 1×10^6 fold (Fig. 1B). If any dCTP is present in the nuclear extract, the remaining concentration after fractionation on DEAE is well below the Kms of the known DNA polymerases. The possibility that dCTP is so tightly bound to the enzyme that it could not be replaced by the exogenous 32 P labeled dCTP is very remote since an enzyme using dCTP as substrate must readily exchange dCTP.

The active fraction 10 was further fractionated sequentially on the following columns: S-Sepharose and Q-Sepharose. The DNA dMTase eluted at the high salt fraction from both columns as determined by the [mdC 32 pdG]n demethylation assay (Fig. 1A). The ion exchange chromatography was followed by chromatography on DEAE-Sephacel.

The fact that we have maintained our activity even after 4 fractionation steps (Table 1) and that only a single polypeptide is apparent after the last purification step argues strongly against the possibility that the activity detected in our study is a repair or replacement activity. Any replacement mechanism must involve a number of proteins and additional cofactors and substrates. In summary, the chromatography of the demethylase activity in A459 cells provides strong

support to the hypothesis that mammalian cells bear a *bona fide* demethylase activity.

DNA dMTase releases a volatile derivative

A *bona fide* demethylation has to result in
5 release of the methyl group as a volatile derivative
such as CO₂, methanol, methane or formaldehyde. We
have therefore incubated a {[³H] -CH₃-dCpdG)n double
stranded oligonucleotide with the different column
fractions and the rate of release of the tritiated
10 methyl from the aqueous phase was determined by scin-
tillation counting of the remaining radioactivity in
the reaction mix. As demonstrated in Fig. 1b (dia-
mond), the dMTase active fractions release labeled
methyl groups from the methylated substrate.

15 DNA dMTase is a protein which is inhibited by RNA, does
not involve an exchange activity and does not require
additional cofactors

DNA dMTase activity measured either as transfor-
20 mation of mdC to C (Fig. 2a) or as release of volatile
methyl residues (Fig. 2c) is abolished after proteinase
K treatment and is not inhibited but rather enhanced
following RNase treatment. 500 μM of ddCTP which
inhibits DNA polymerase does not inhibit demeth-
25 ylation of the [mdC32pdG]n substrate, nor is it inhib-
ited by high concentrations of methyl-dCTP (500 μM)
(Fig. 2a), which is consistent with the hypothesis that
demethylation does not involve an excision and replace-
ment mechanism. If a replacement mechanism is involved
30 in demethylation, the presence of mdCTP should result
in incorporation of methylated cytosines and essential
inhibition of demethylation. Thus, the DNA dMTase
identified here is a protein and not an RNA and is une-
quivocally different from the previously published RNA
35 based or glycosylase based demethylase activities.

The DNA dMTase reaction proceeds without any requirement for additional substrates such as dCTP, redox factors such as NADH and NADPH or energy sources such as ATP (data not shown). As observed in Fig. 2b and 2d, the DNA dMTase reaction maintains its initial velocity up to 90 minutes and continues up to 120 minutes. This time course is inconsistent with dependence on enzyme-bound additional nonreplenishable substrates such as dCTP or ATP or a nonreplenishable redox factor such as NADH or NADPH. Exhausting the nonreplenishable substrate or redox factor would have resulted in rapid deceleration of the initial velocity.

A product of the demethylation reaction is deoxyCytosine in DNA

What is the product of the demethylation reaction? The results presented above (Fig.1a, 2a and b) based on a one dimension TLC separation show that DNA dMTase generates dC from mdC in DNA. To further substantiate this conclusion, we subjected DNA dMTase treated DNA to remethylation with the CpG MeTase M.Sss I which can transfer a methyl group exclusively to dC. The results presented in Fig. 3a show that the demethylated product of DNA dMTase is dC since it is completely remethylated with M.Sss I. The identity of the demethylated product as dC was further established by a two-dimension TLC analysis demonstrating that the product of dMTase comigrates with a cold dCMP standard in both dimensions (Fig. 3b).

DNA dMTase does not release a nucleotide, a phosphorylated base or phosphate from methylated DNA when incubated with a [32 PmdCpdG] $_n$ substrate which included a labeled 32P 5' to mdC or our standard methylated substrate (Fig.1) where 32P is 3' to the m5dC (Fig. 3c). Nuclear extracts which obviously contain a number of glycosylases and nucleases release phospho-

rylated derivatives in the same assay (Fig. 3c). dMTase transforms the methyl cytosine in the [32pmdCpdG]n substrate to cytosine as demonstrated when the reacted DNA is digested to 5' mononucleotides
5 (Fig. 3c +V PDS) and analyzed by TLC. Since this reaction does not involve release of a 32P derivative (Fig. 3c -V PDS), it demonstrates that dMTase transforms methylated cytosines to cytosines on DNA without disrupting the integrity of the DNA substrate by glyco-
10 sylase or nuclease activity .

The second product of the dMTase reaction is methanol

What is the identity of the leaving group? The results presented in Fig1b suggest that the labeled
15 methyl leaves the DNA as a volatile compound. The demethylase reaction involves release of the methyl group *per se* whereas the cytosine base ring remains in the aqueous phase. Fig. 4a demonstrates this point by using a methylated plasmid labeled with a ³H-hydrogen
20 at the sixth position of cytosine and [14C]-methyl at the fifth position of cytosine as a substrate.

The three most obvious candidates the methyl group is leaving as are formaldehyde, carbon dioxide, and methanol. Methadone trapping for labeled formalde-
25 hyde detection and sodium hydroxide trapping for labeled carbon dioxide detection were both negative in identifying the form in which the methyl group is leaving in the dMTase reaction (data not shown). The other possible chemical form that the methyl group may leave
30 the DNA as, is methanol. Since methanol is a volatile compound, a simple method to measure generation of methanol is a scintillation-volatilization assay (see Fig. 4b for description). Volatilization assays have been previously used to measure release of methanol in
35 demethylation reactions. The demethylation reaction mix containing the labeled {[³H] -CH₃-dCpdG}n substrate

with either dMTase or no enzyme, as a control, is added to an uncapped 0.5 ml tube which is placed in a sealed scintillation vial containing scintillation fluid. Released methanol is volatile, diffuses out of the open
5 reaction tube and is mixed with the excess of the scintillation fluid in the vial registering as counts in the scintillation counter. As a control indicating that methanol is volatilized under the conditions of our assay, we incubated approximately equal counts of
10 radioactively labeled methanol under the same conditions and measured the counts in a scintillation counter at different time points. As observed in Fig. 4c the majority of methanol in the reaction tube volatilizes from the reaction tube into the scintillation
15 fluid following an overnight incubation at 37°C. The experiment shown in Fig. 4b demonstrates that volatilized label is released from methylated DNA only in the presence of dMTase.

The identity of the volatile group has been
20 determined to be methanol by a gas chromatography (GC) analysis. The demethylation and control reactions (indicated in Fig. 4e) were performed in an uncapped tube placed in a sealed scintillation vial containing a larger volume (300 μ l) of water. The volatile residue
25 diffuses into the surrounding water and mixes with it. A 2 μ l sample of the surrounding water was injected into a GC column as described in the methods. As shown in Fig. 4e, the volatile compound released by dMTase in a dose response manner coelutes with metha-
30 nol. Release of methanol is observed only in the presence of both dMTase and methylated DNA. No methanol is released when dMTase is reacted with nonmethylated DNA, demonstrating that methanol is a product of demethylation of DNA.

The leaving group was also identified as methanol using gas chromatography coupled with Mass spectrometry (GC-MS). As illustrated in Fig. 4f., incubation of methylated DNA with dMTase (dMTase+ME-DNA) results in release of a peak with the retention time and mass spectrum (peaks are identified at 32 and 29 atomic mass which are the atomic masses of methanol and ionized methanol respectively) which is consistent with its identification as methanol. Incubation of dMTase with nonmethylated DNA does not release methanol indicating that methanol is a product of the demethylation reaction. No methanol is released when the samples are incubated with dMTase treated with protease K indicating that the release of methanol from methylated DNA is catalyzed by an enzymatic activity.

Demethylation involves transfer of a hydrogen from water to regenerate cytosine

If demethylation involves removal of the methyl moiety from mdC, a hydrogen has to be transferred to the carbon at the 5' position to regenerate cytosine. Since no redox factors are involved, what is the source of the hydrogen? To test the hypothesis that the source of the hydrogen is water, we incubated either non labeled [mdCpdG]n or [dCpdG]n double stranded DNA with DNA dMTase for different time periods in the presence of tritiated water, following which the DNAs were digested to 3' dNMPs, separated on TLC with non-radioactive standards for each of the 5 possible dNMPs and exposed to a tritium sensitive phosphorimaging plate. As seen in Fig.4d, dMTase catalyzes the transfer of a tritiated hydrogen from water to dCMP in methylated DNA in a time dependent manner only when methylated DNA is used as a substrate. Based on the experiments described in Fig.3 and 4 we propose that dMTase catalyzes the exchange of the methyl group at

the 5' position of cytosine in DNA with hydrogen from water and the methyl group reacts with the remaining hydroxyl group to form methanol (Fig. 5).

5 Substrate and sequence specificity of DNA dMTase

Methylation of CpG dinucleotides is the most characterized modification occurring in genomic DNA^{8,48}. The results presented in Fig.6 demonstrate that DNA dMTase is a general DNA dMTase activity that demethylates fully or hemimethylated dCpdG in DNA flanked by a variety of sequences which are distributed at different frequencies, but does not demethylate methylated adenines or methylated cytosines that do not reside in the dinucleotide CG. First, as shown in Fig.6a, a plasmid DNA methylated *in vitro* at all dCpdG sites with M.Sss I and all d*CdCdGdG sites with M. Msp I (which methylates the external C in the sequence *CCGG, thus enabling the determination of demethylation at the CC dinucleotide) and *in vivo* with the *E. coli* DCM MeTase at dCmdCdA/dTdGdG sites and with the DAM MeTase at dGmdAdTdC sites (adenine methylated) was treated with dMTase and the state of methylation of the plasmid was determined using the indicated methylation sensitive restriction enzymes. dMTase demethylates C*G methylated sites as indicated by the sensitivity of the dMTase treated plasmid to *Hpa* II and *Hha* I but does not demethylate C*C,C*A or C*T methylated sites as indicated by the resistance to *Msp* I and *Eco* RII restriction enzymes, or adenine methylation as indicated by its sensitivity to *Dpn* I. Second, bisulfite mapping analysis of methylation of 5 methylated C*G sites residing in a M.Sss I *in vitro* methylated pMetCAT plasmid following dMTase treatment shows that all C*G sites are demethylated irrespective of their flanking sequences thus excluding the possibility that demethylation is limited to CCGG or CGCG sequences (Fig. 6b).

Third, dMTase does not demethylate two fully methylated cytosine bearing oligomers [dmC32pdA]_n, [mdC32pdT]_n demonstrating that mdCpdA and mdCpdT are not demethylated by DNA dMTase (Fig. 6d). Fourth, dMTase demethylates a hemimethylated synthetic substrate [dCpdG]_n*[mdC32pdG]_n (Fig. 6d). Demethylation of SK is complete under these conditions (Fig. 6a) whereas demethylation of a methylated [mdCpdG]_n substrate is not complete under the same conditions (Fig. 6d). This can reflect differences in the sequence composition of the substrate and the frequency of methylated cytosines. The [mdCpdG]_n contains on average 16 fold more methylated cytosines per molecule than plasmid DNA. Alternatively, these differences might reflect discrepancies in the assays used, restriction enzyme digestion versus a nearest neighbor analysis. To address this discrepancy we have labeled a fully methylated SK plasmid with [α^{32} P]dCTP, 5-methyl-dCTP and the other dNTPs, subjected it to dMTase treatment and digested it to mononucleotides at different time points following the initiation of the reaction and subjected the samples to a TLC analysis. As shown in Fig. 6c, the SK plasmid is fully demethylated at 3 hours which is consistent with the results obtained with methylation sensitive restriction enzymes (Fig. 6a).

The K_m of DNA dMTase for hemimethylated and fully methylated DNA was determined by measuring the initial velocity of the reaction at different concentrations of substrate (Table 2). The calculated K_m for hemimethylated DNA is 6 nM which is two fold higher than the K_m for DNA methylated on both strands, 2.5-3 nM (Table 2). It is unclear yet whether this small difference in affinity to the substrate has any significance in a cellular context. Thus similar to the DNA MeTase DNA dMTase shows dinucleotide sequence

selectivity but in difference from DNA MeTase which shows preference to hemimethylated substrates dMTase prefers fully methylated DNA which is consistent with a role for DNA dMTase in altering established methylation patterns.

Table 1

Purification of DNA dMTase

| Purification step | Total protein (μ g) | Total dpm | pMole/ μ g | pMole/ μ g/h | Fold Purification |
|-------------------|--------------------------|-----------|----------------------|------------------------|-------------------|
| Nuclear extract | 6000 | 1107.2 | 5.5×10^{-5} | 1.833×10^{-5} | - |
| DEAE-Sephadex | 3.75 | 5844 | 0.4674 | 0.156 | 8445.5 |
| SP-Sepharose | 0.77 | 5106 | 1.989 | 0.663 | 35939.84 |
| Q-Sepharose | 0.46 | 5335 | 3.4 | 1.13 | 62860.65 |
| DEAE-Sepharcel | 0.018 | 1834 | 30.57 | 10.19 | 552243.2 |

Table 2

Kinetic parameters for DNA dMTase

| Method | K_m (DNA) | V_{max} (pMole/h) |
|----------------------|-------------|---------------------|
| Methylated oligo CpG | 2.5 nM | 340 |
| Hemi-methylated CpG | 6.0 nM | 402 |
| Methylated SK-DNA | 3.3 nM | 40.42 |

Cloning and construction of demethylase expression vectors

15 PCR amplification of the MBD domain of the putative demethylase candidate cDNA

One μ g of total RNA prepared from the human small lung carcinoma cell line A549 was reverse transcribed using Superscript reverse transcriptase and random primers (Boehringer) in a 25 μ l reaction volume according to conditions recommended by the manufacturer (GIBCO-BRL). Five μ l of reverse transcribed cDNA were subjected to an amplification reaction with Taq polymerase (Promega, 1 unit) using the following set of

primers: sense 5'CTGGCAAGAGCGATGTC 3' SEQ ID NO:9,
antisense 5'AGTCTGGTTTACCCTTATTTTG 3' SEQ ID NO:10.

Amplification conditions were: step 1. 95°C 1 min.; step 2: 94°C 0.5 min; step 3: 45°C 0.5 min.; step
5 4: 72°C 1.5 min; steps 2-4 were repeated 30 times. MgCl₂ was adjusted to 1 mM according to conditions recommended by the manufacturer. The PCR products were cloned in pCR2.1 vector (Invitrogen) and the sequence of the cDNAs was verified by dideoxy-chain termination
10 method using a T7 DNA sequencing kit (Pharmacia). The amplified fragment was excised from the plasmid with EcoRI, labeled with a Boehringer random prime labeling kit according to manufacturer's protocol and alpha ³²P-dCTP. The labeled probe was used to screen a HeLa cell
15 cDNA library in λTriplEx phage (Clontech) according to standard procedures. Positive clones were identified and further purified by serial dilutions for 4 rounds. The insert in the pTriplEx plasmid was excised from the phage according to manufacturer's protocols and the
20 identity of the insert was verified by sequencing. The insert was excised by NotI restriction and subcloned into either the inducible expression vector: Retro tet on (Clontech) in the sense and antisense orientation or the pcDNA3.1/His Xpress vector in all three frames and
25 in the antisense orientation.

Transfection and expression of demethylase in vertebrate cells

Ten µg of either Retro tet on demethylase or
30 pcDNA 3.1/His Xpress demethylase are mixed with 8 µl of transfection lypophilic reagent Pfx-2 (Invitrogen) and placed upon 100,000 mouse (3T3 Balb/c, human (A549) or monkey cells (CV-1) according to manufacturer's protocol in OPTIMEM medium for 4 hours. Cells are harvested
35 after 48 hours and demethylation and demethylase activity is determined by measuring total genomic DNA meth-

ylation using standard techniques or a cotransfected *in vitro* methylated plasmid using a HpaII /MspI restriction enzyme analysis. Cellular transformation is measured by a soft agar assay.

5

Demethylation of pBluescript SK(+) Plasmid

About 4 μ g plasmid pBluescript SK (Stratagene) was subjected to methylation using SssI methylase. The methylated plasmid (4 ng) was incubated for different
10 time points as indicated with 30 μ l of DNA dMTase Fraction 4 of DEAE-SephacelTM column under standard conditions, extracted with phenol: chloroform and precipitated with ethanol. About 1 ng of the plasmid were subjected to digestion with 10 units each of either of
15 the restriction endonuclease EcoRII (GIBCO-BRL), DpnI, or HpaII (New England Biolabs) before and after methylation as well as after DNA dMTase treatment in a reaction volume of 10 μ l for 2 hour at 37°C. Following restriction digestion the plasmids were extracted with
20 phenol:chloroform, ethanol precipitated and resuspended in 10 μ l. The plasmids were electrophoresed on a 0.8% (w/w) Agarose gel, transferred onto a HybondTM Nylon membrane and hybridized with pBluescript SK(+) plasmid which was ³²P labeled by random-priming (Boehringer
25 Mannheim).

dMTase activity coelutes with a ~45 KDa polypeptide when sized under denaturing conditions but migrates as a higher molecular weight complex under non denaturing
30 conditions. dMTase was purified up to 500,000 fold by four chromatographic steps (Table 1). We first determined the identity of the polypeptide associated with dMTase activity by SDS-PAGE analysis of the active fractions. As observed in Fig. 7a, a cluster of 4
35 polypeptide bands from ~44 KDa to 35 KDa coelute with dMTase activity in the last two chromatographic steps

(the lower fragment might be a degradation product as evidenced by its abundance in the later chromatographic steps). However when the active DEAE-Sephacel fraction is size fractionated on a 4% non denaturing acrylamide column, the dMTase activity elutes at the high molecular weight of ~170 KDa (Fig. 7c, fraction 63). SDS-PAGE analysis of this fraction (63) reveals only two bands (Fig. 7b) observed in the active chromatographic fractions (Fig. 7a). To further determine whether dMTase is found in a multimeric complex, fraction 63 was size fractionated on a glycerol gradient (Fig. 7d) and DNA dMTase activity eluted at the ~170 kDa range. As only two main small polypeptides were identified in fraction 63 (approximately 35-43 KDa), dMTase is probably found in either a homomeric complex if only one of the two peptides is dMTase or a heteromeric complex if both polypeptides are associated with dMTase activity.

a. Identification of a lead DNA dMTase candidate by homology search of dbEST

As the purification of dMTase suggests that the dMTase is of very low abundance, only ~19 ng of dMTase could be isolated from 6 mg of nuclear extract (Table 1), we opted for cloning the dMTase based on its following functional properties. First, since dMTase specifically demethylates methylated CG dinucleotides, we assumed that it should bear the ability to recognize methylated CG dinucleotides. Second, the demethylase transforms methylated cytosine in DNA to cytosine. Third, the demethylase releases the methyl group as a volatile compound.

Previous reports have shown that proteins interacting with methylated DNA share a common domain (MDBD). A TBLASTN search of the dbEST database identified a novel expression tag cDNA (from a T-cell lymphoma Homo sapiens cDNA 5' end) (gb/AA361957/AA361957

EST71295) and the mouse homologue ((gb/W97165/W97165
mf90g05.r1) from Soares mouse embryo NbME13.5) with
unknown function that bears homology to the MDBD
(Fig. 8a). A search of the GenBank database verified
5 that it is a novel cDNA that has not been included in
GenBank. Alignment of the novel EST and MeCP2 and
MeCP1 associated protein has revealed no homology
beyond the previously characterized MDBD which is con-
sistent with a different function for this methylated
10 DNA binding protein. A 201bp fragment bearing the
sequence identified in the search was reverse tran-
scribed and amplified from human lung cancer cell line
A549 RNA and was used to screen a cDNA library from
Hela cells. The largest insert cloned was of 1.36 kb
15 size and its sequence identity with the EST sequence
was determined. The cDNA is novel and has no homologue
in GenBank and no function has ever been assigned to
it. A virtual translation of the protein identified an
open reading frame (ORF) of 262 amino acids (Fig. 8b).
20 The ORF may extend further 5' as no in frame stop codon
was found upstream of this ATG. However, RACE analy-
ses and further searches of the dbEST have failed to
identify 5' sequences upstream to the one identified in
our screening.

25 A BLAST search of the candidate protein using
the Predict protein server against a database of pro-
tein domain families has identified only the MDBD
domain and found no homologue to the sequence in the
data base search. No other functional motifs were
30 identified by the Prosite analysis. This is consistent
with a novel biochemical function for this protein. A
coiled coil prediction of the sequence identified a
coiled coil domain which is known to play a role in
protein protein interactions.

The identified cDNA encodes an mRNA that is widely expressed in human cells as revealed by a Northern blot analysis of human poly A+ mRNA (Fig. 8c) as one major transcript of ~ 1.6 kb which is close to the size of the cloned cDNA, verifying that the cloned cDNA does not represent a highly repetitive RNA but rather a mRNA encoded by a single or low copy number gene.

In vitro translated candidate cDNA bears dMTase activity

A conclusive proof for the existence of a single protein that *bona fide* demethylates DNA is to demonstrate that an *in vitro* translated candidate cDNA can volatilize methyl groups from methylated DNA and transform a methyl cytosine to cytosine in an isolated system. The candidate dMTase cDNA was subcloned it into a pcDNA3.1/His Xpress (INVITROGEN) expression vector in the putative translation frame (pcDNA3.1His A) and in a single base frame shift (pcDNA3.1His B), and was *in vitro* transcribed and translated in the presence of ³⁵S-methionine and the resulting translation products were resolved by SDS-PAGE. Autoradiography revealed a ~40KDa protein (Fig. 10a). The apparent size of the *in vitro* translated protein is shorter by ~3-5 KDa from the apparent size of the purified protein. The cloned cDNA might be missing some upstream amino acids as discussed above or might be differently modified in human cells.

Two tests established whether the *in vitro* translated candidate cDNA is a *bona fide* dMTase. We first tested whether *in vitro* translated protein (purified on a Ni²⁺ charged agarose resin) can volatilize and release methyl residues in [³H]-CH₃-DNA using a radioactive trapping volatilization assay. To verify that the volatilized counts are true ³H counts, a spectrum analysis was performed. As demonstrated in Fig.

10b no volatilization of tritiated methyl residues is observed in the misframe dMTase (misframe) whereas *in vitro* translated putative dMTase cDNA catalyzes the volatilization of $^3\text{H-CH}_3$ residues which are trapped in the scintillation cocktail.

Second, *in vitro* translated dMTase cDNA transforms CH_3 -cytosine residing in $[\text{P}^{32}]\text{-}\alpha\text{-dGTP}$ labeled plasmid DNA or in [methyl-dC32pdG]n double stranded oligomer DNA to cytosine, whereas a frame shift *in vitro* translated dMTase does not demethylate DNA (Fig. 10d). This demonstrates that the dMTase activity is dependent on the dMTase translation product and not a contaminating activity found in the *in vitro* translation kit that copurifies with the putative dMTase. The reaction carried out by the *in vitro* translated dMTase displays: dependence on the dose of *in vitro* translated product (Fig. 10c), time dependence (Fig. 10d) and dependence on translated protein (Fig. 10b & d misframe, Fig. 10c protease K treatment). Taken together, these results strongly suggest that the cDNA cloned here codes for a *bona fide* enzymatic DNA demethylase activity.

Transiently transfected dMTase cDNA demethylates DNA

dMTase cDNA and the pcDNA3.1HisC vector control were transiently transfected into human embryonal kidney cells to test whether the cDNA can direct expression of dMTase activity in human cells. The His-tagged proteins were bound to Ni^{2+} agarose resin and eluted from the resin with increasing concentrations of imidazole. The expression of the transfected dMTase was verified by a Western blot analysis (Fig. 11b). The imidazole fractions were assayed for their ability to volatilize and release methyl residues in $[\text{H}^3]\text{-CH}_3\text{-DNA}$ using a radioactive trapping volatilization assay 1. As observed in Fig. 11a, imidazole fractions from

dMTase transfected cells volatilize [^3H]-CH₃, whereas no tritiated counts are detected in DNA treated with imidazole fractions from cells transfected with a misframe mutation of dMTase or non transfected cells. The transiently expressed dMTase transforms methylated cytosine in DNA to cytosine residing in two different substrates (Figs. 11c & 11d), in a protein dependent manner (Figs. 11c & 11e), and the reaction displays substrate dependence and saturability (Fig. 11f). Transiently expressed dMTase was loaded on a non denaturing glycerol gradient to determine its native MW. Similar to dMTase purified from human cells, cloned and purified dMTase activity fractionated at the 160-190 KDa range (data not shown). This is consistent with self association of cloned dMTase possibly mediated by the coiled-coil domain.

Cloned DNA dMTase catalyzes a hydrolysis of 5-methylcytosine to release methanol

We determined the mechanism by which methyl residues are released by the cloned dMTase (from Fig. 11) and compared it to the purified *bona fide* dMTase activity. Increasing amounts of non labeled [methyl-dCpdG] DNA were incubated with either the *bona fide* dMTase activity purified from A549 cells or the cloned dMTase in the presence of [^3H] water for 3 hours followed by digestion to mononucleotides, a thin layer chromatography and autoradiography. As Fig. 12a shows, both reactions replace the methyl group in 5-methylcytosine with a proton donated from water as indicated by the presence of [^3H] label in cytosine.

The identity of the leaving methyl group in the demethylation reaction catalyzed by the purified *bona fide* dMTase activity was shown to be methanol. In order to identify the form that the methyl residue leaves as in the demethylation reaction catalyzed by

the cloned dMTase an identical gas chromatography/mass spectrometry analysis of the reaction products was performed as in1. Only the properly translated form of dMTase (both in vitro translated and transiently transfected and purified) is able to produce ions characteristic of methanol in a mass spectrometric analysis (mass of 32 and 29, Fig. 12b). These results suggest that the demethylation reaction catalyzed by the cloned dMTase is hydrolysis of the 5-methyl-cytosine to cytosine and methanol as described for the purified dMTase1.

DNA dMTase activity is undetectable in nontransformed cells

The assays for dMTase activity described here and the cloning of DNA dMTase cDNA enables a study of its expression at different cellular states. Global hypomethylation of DNA is a common observation in cancer cells. This has been a perplexing observation, since DNA MeTase activity is elevated in cancer cells. Hyperactivation of DNA MeTase has been proposed to play a role in cancer development. This paradox raises questions on the proposed role of the elevated levels of DNA MeTase in cancer cells. One simple explanation that has been previously suggested to resolve this paradox is that cancer cells express induced levels of DNA dMTase. We compared the DNA dMTase activity in equal concentrations of DEAE-Sephadex fractionated nuclear extracts (fractions 9-10) prepared from a number of carcinoma cell lines H446, Colo 205, Hela, and A549 with a similar preparation from human skin fibroblast cells at initial rate conditions using [mdC32pdG]n double stranded oligomer as a substrate. As observed in Fig. 13a, whereas DNA dMTase activity is readily observed in all carcinoma cell lines, it is undetectable in nontransformed human cells. The absence of dMTase activity in human primary cells

reflects the situation *in vivo* since dMTase activity is undetectable in preparations from different murine tissues whereas dMTase activity is present in a murine carcinoma cell line P19 that was transfected with the

5 H-Ras protooncogene, or human tumors carried as xenografts in the same strain of mouse (Fig. 1a: COLO 205, A549, Hela). These conclusions were verified using the radioactive-trapping volatilization assay shown in Fig. 13c.

10 Since dMTase mRNA has been detected using a sensitive poly A+ Northern blot in all normal human tissues, we tested the hypothesis that the absence of detected dMTase activity in normal tissues reflects a quantitative difference in DNA dMTase mRNA between normal

15 tissues and cancer lines. A Northern blot analysis and quantification of dMTase mRNA by a slot blot analysis shown in Fig. 13d using total RNA supports this hypothesis. Whereas minute levels of dMTase mRNA are detected in normal tissues, high levels of dMTase are

20 expressed in a murine carcinoma cell line Y1 that bears a 30 fold amplification of *Ha-ras*.

A second DNA demethylase dMTase2 identified in human and mouse

cDNA sequences, predicted amino acid sequences, and

25 GenBank accession numbers of both dMTase1 and dMTase2 from human and mouse are shown. We claim that the high level of identity of the two proteins (Figs 9c and e) suggests that the two proteins can perform the same function, DNA demethylation. The N-terminals of

30 dMTase1 and dMTase2 contain a Methylated DNA Binding Domain (MBD) and near their C-terminals is a coiled-coil domain, however the middle portions of the protein sequences have no homology to any known structural or catalytic motif. Importantly, their middle regions are

35 still extensively homologous suggesting that the cata-

lytic site of the demethylase activity lies in this area on both proteins.

Induced expression of DNA demethylase in the Antisense orientation inhibits tumorigenesis *ex vivo*

5 To test the hypothesis that inhibition of DNA dMTase can inhibit tumorigenesis tetracycline inducible vectors carrying the human dMTase1 cDNA in either the sense or antisense orientation were constructed and transiently transfected into HEK 293 cells, treated for
10 48 hours either in the presence or absence of doxycycline (a tetracycline analogue), selected for the last 24 hours with puromycin, and then plated on soft agar and allowed to grow for seven days. After seven days colonies were scored and the data presented clearly
15 show that doxycycline induced expression of the dMTase1 cDNA in the antisense orientation reduced colony formation (Fig. 15).

Imidazole is a small molecule inhibitor of DNA demethylase activity

20 A template small molecule, imidazole, was tested for the ability to inhibit DNA dMTase activity. In a volatilization of radioactive methyl residues assay, concentrations from 1 μ M to 10mM of imidazole were incubated in a typical volatilization of radioactive methyl
25 residues as described above. The graph clearly demonstrates a dose dependent inhibition of DNA dMTase activity by imidazole, and validates a rationale for testing imidazole based molecules as inhibitors of DNA dMTase activity (Fig. 16).

30 Identification of DNA demethylase cDNAs and protein sequences

Fig. 9a illustrates cDNA sequence of human dMTase1 (SEQ ID NO:1) and its predicted amino acid sequence (SEQ ID NO:2), including its Genbank location. Fig. 9b illustrates cDNA sequence of human dMTase2 (SEQ ID NO:3) and
35 its predicted amino acid sequence (SEQ ID NO:4), includ-

ing its GenBank location. Fig. 9c illustrates protein sequence alignment of human dMTase1 and human dMTase2. Fig. 9d illustrates cDNA sequence of mouse dMTase1 (SEQ ID NO:5) and its predicted amino acid sequence (SEQ ID NO:6), including its GenBank location. Fig. 9e illustrates cDNA sequence of mouse dMTase2 (SEQ ID NO:7) and its predicted amino acid sequence (SEQ ID NO:8), including its GenBank location. Fig. 9f illustrates protein sequence alignment of mouse dMTase1 and mouse dMTase2.

While the invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications and this application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure as come within known or customary practice within the art to which the invention pertains and as may be applied to the essential features hereinbefore set forth, and as follows in the scope of the appended claims.

WHAT IS CLAIMED IS:

1. A DNA demethylase enzyme and/or homologue thereof having about 40 KDa, and wherein said DNA demethylase enzyme is overexpressed in cancer cells.
2. A cDNA encoding a human demethylase which comprises a sequence set forth in SEQ ID NOS:1 and 3.
3. A cDNA homologous to the cDNA of claim 2, wherein said cDNA encoding mouse demethylase set forth in SEQ ID NOS:5 and 7.
4. The use of the expression of demethylase cDNA of claims 2 or 3 to alter DNA methylation patterns of DNA *in vitro* in cells or *in vivo* in humans, animals and in plants.
5. The use of claim 4, wherein said demethylase cDNA expression is under the direction of mammalian promoters.
6. The use of claim 5, wherein said promoter is CMV.
7. The use of claim 4, wherein said demethylase cDNA expression is under plant specific promoters to alter methylation in plants and to allow for altering states of development of plants and expression of foreign genes in plants.
8. The use of claim 4, wherein said demethylase cDNA expression is in the antisense orientation to inhibit demethylase in cancer cells for therapeutic processes.

9. The use of claim 9, wherein expression of demethylase cDNA in mammalian cells is to alter their differentiation state and to generate stem cells for therapeutics, cells for animal cloning and to improve expression of foreign genes.
10. The use of the expression of demethylase cDNA of claims 2 or 3 in bacterial or insect cells for production of large amounts of demethylase.
11. The use of the expression of demethylase cDNA of claims 2 or 3 for the production of protein in vertebrate, insect or bacterial cells.
12. The use of claim 11 for producing antibodies against demethylase.
13. The use of the sequence of demethylase cDNA of claim 2 as a template to design antisense oligonucleotides and ribozymes.
14. The use of the predicted peptide sequence of demethylase cDNA of claim 2 to produce polyclonal or monoclonal antibodies against demethylase.
15. The use of expression of cDNA of claim 2 or 3 in two hybrid systems in yeast to identify proteins interacting with demethylase for diagnostic and therapeutic purposes.
16. The use of expression of cDNA of claim 2 or 3 in bacterial, vertebrate or insect cells to produce large amounts of demethylase for high throughput screening of

demethylase inhibitors for therapeutics and biotechnology and for obtaining the x-ray crystal structure.

17. A volatile assay for high throughput screening of demethylase inhibitors as therapeutics and anticancer agents which comprises the steps of:

- a) using transcribed and translated demethylase cDNA of claim 2 or 3 in vitro to convert methylcytosine present in methylated DNA samples to cytosine present in DNA and volatilize methyl group;
- b) determining the absence or minute amount of volatilized methyl group as an indication of an active demethylase inhibitor.

18. A volatile assay for the diagnostics of cancer in a patient sample which comprises the steps of:

- a) determining demethylase activity in patient samples by determining conversion of methylcytosine present in methylated DNA to cytosine present in DNA and volatilization of the methyl group released as methanol;
- b) determining the presence or minute amount of volatilized methyl group as an indication of cancer in said patient sample.

19. Use of an antagonist or inhibitor of DNA demethylase of claim 1 or 2 for the manufacture of a medication for cancer treatment, for restoring an aberrant methylation pattern in a patient DNA, or for changing a methylation pattern in a patient DNA.

20. Use according to claim 19, wherein said antagonist is a double stranded oligonucleotide that inhibits demethylase at a K_i of 50nM.

21. Use according to claim 20, wherein said oligonucleotide is
$$\begin{array}{c} [C^mGC^mGC^mGC^mG] \\ [G^mCG^mCG^mCG^mC]_n \end{array} .$$

22. Use according to claim 19, wherein the inhibitor comprises an anti-DNA demethylase antibody or an antisense oligonucleotide of DNA demethylase or a small molecule.

23. Use according to one of claims 19 or 22, wherein the change of the methylation pattern activates a silent gene.

24. Use according to claim 23, wherein the activation of a silent gene permits the correction of genetic defect.

25. Use according to claim 24, wherein said genetic defect is β -thalassemia or sickle cell anemia.

26. Use of the demethylase of claim 1, for removing methyl groups on DNA *in vitro*.

27. Use of the demethylase of claim 1 or its cDNA of claim 2, for changing the state of differentiation of a cell to allow gene therapy, stem cell selection or cell cloning.

28. Use of the demethylase of claim 1 or its cDNA, of claim 2 for inhibiting methylation in cancer cells using vector mediated gene therapy.

29. An assay for the diagnostic of cancer in a patient, which comprises determining the level of expression of DNA demethylase of claim 1 in a sample

from said patient, wherein overexpression of said DNA demethylase is indicative of cancer cells.

1/50

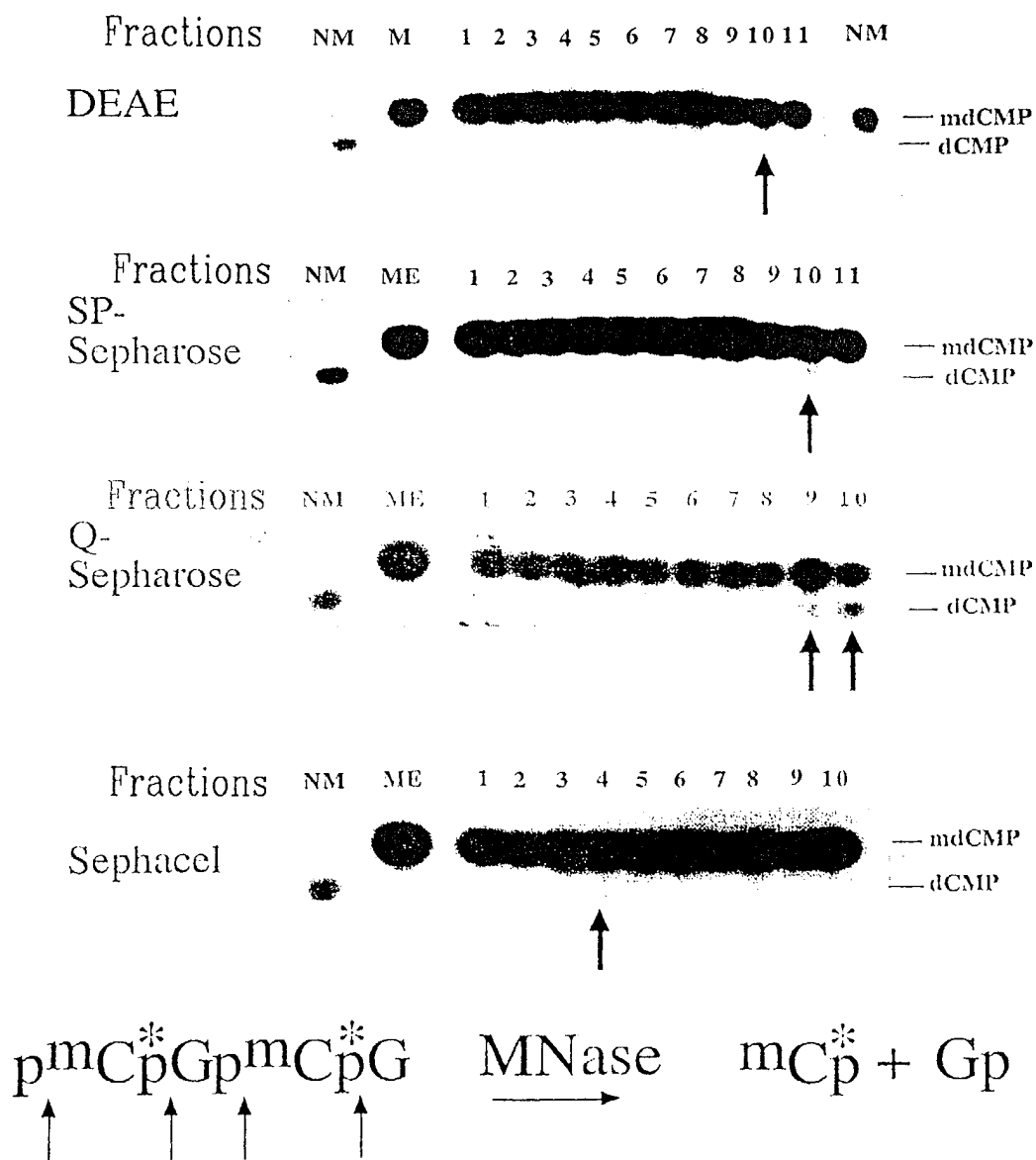


Fig. 1A

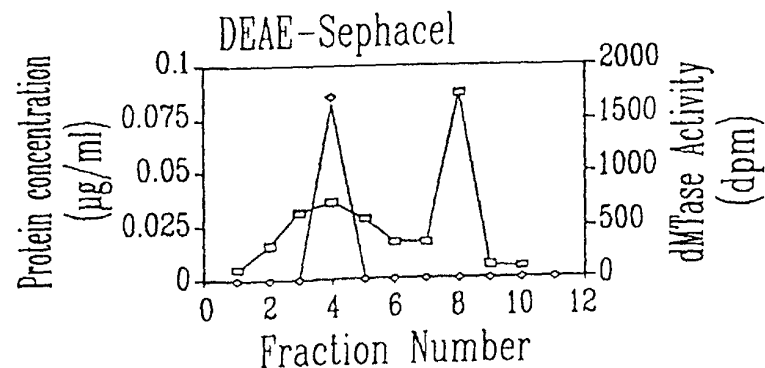
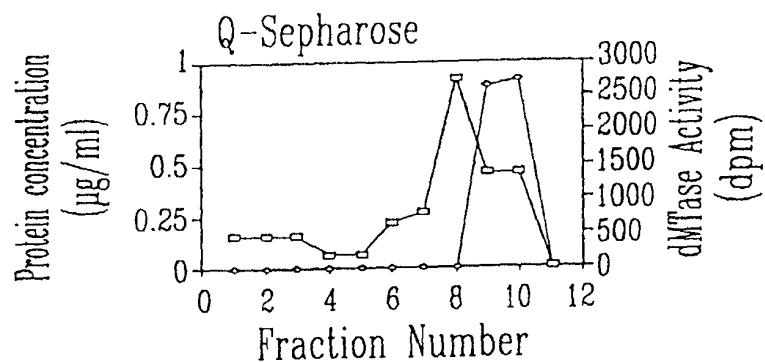
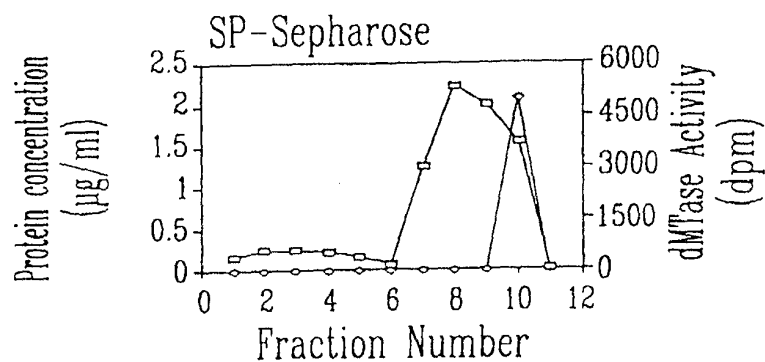
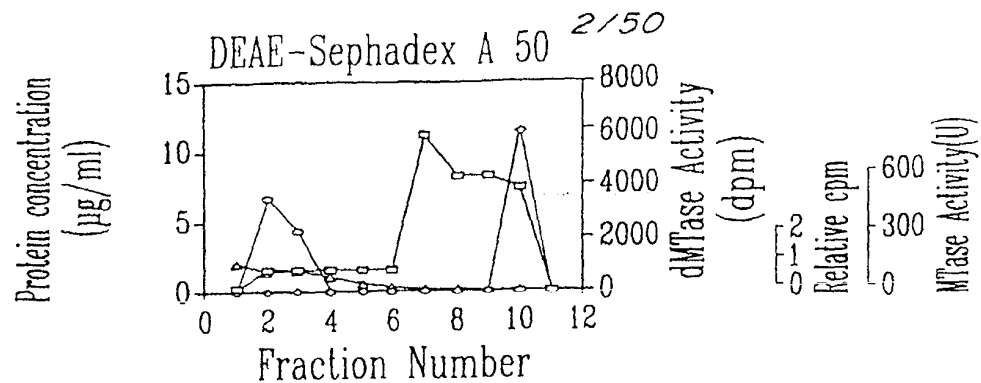
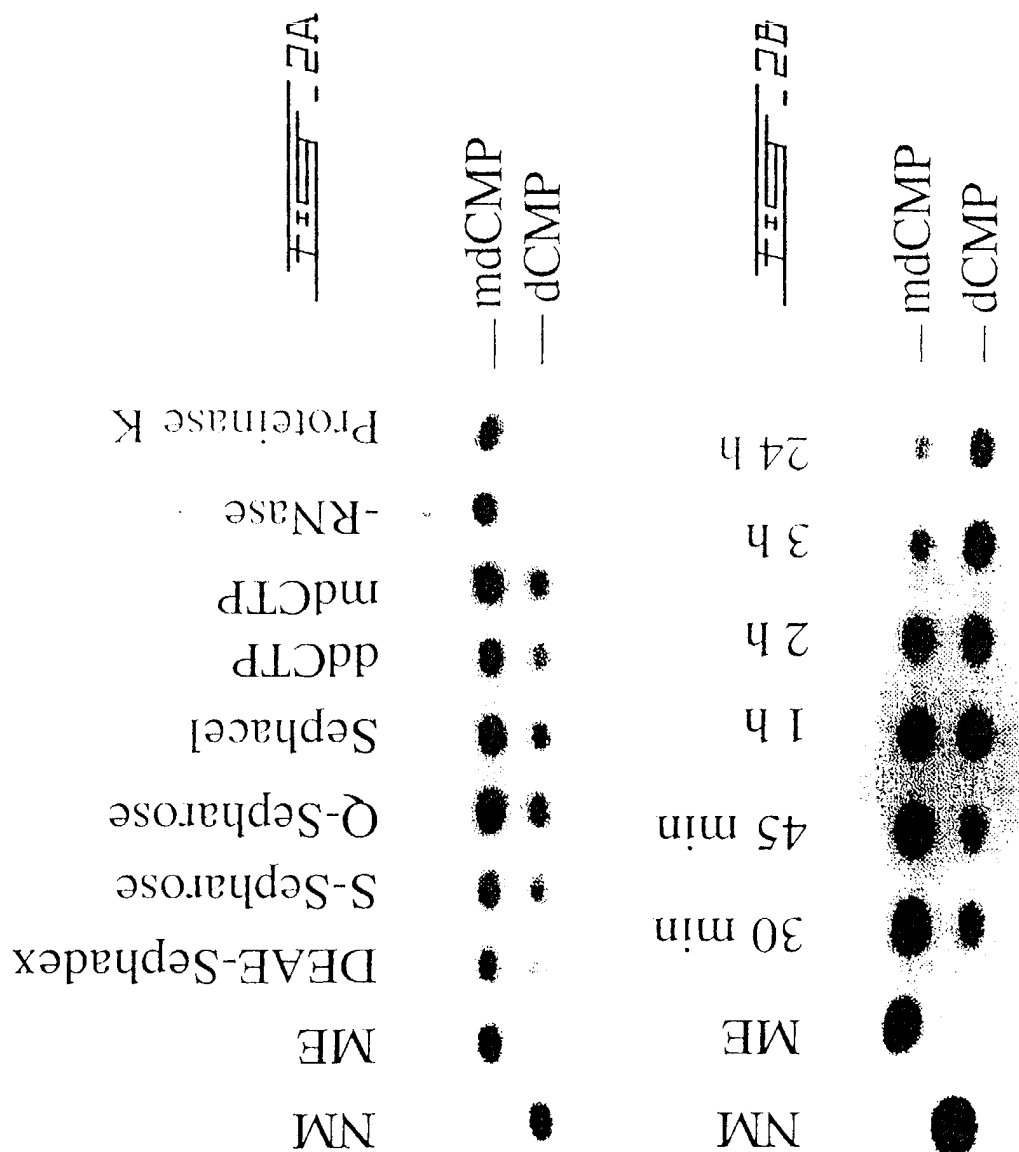


FIG. 1B

3/50



4/50

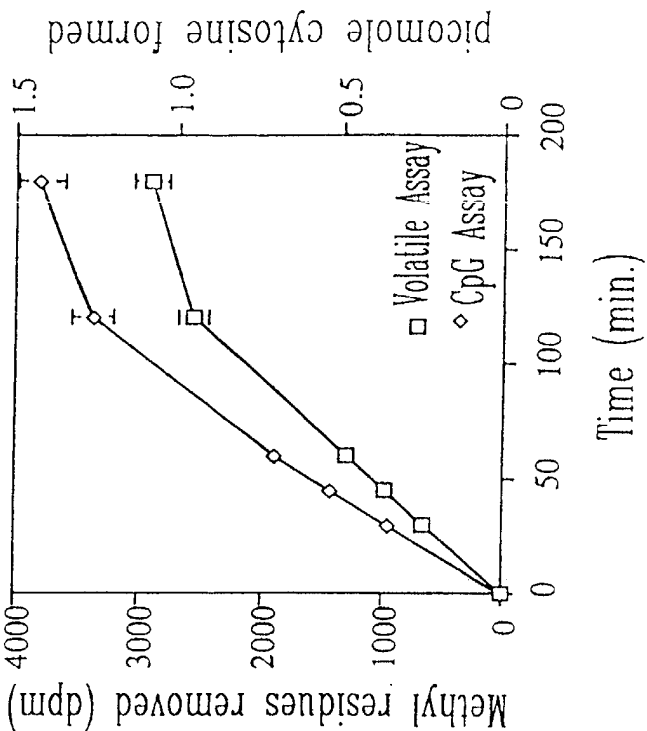


FIG. 20

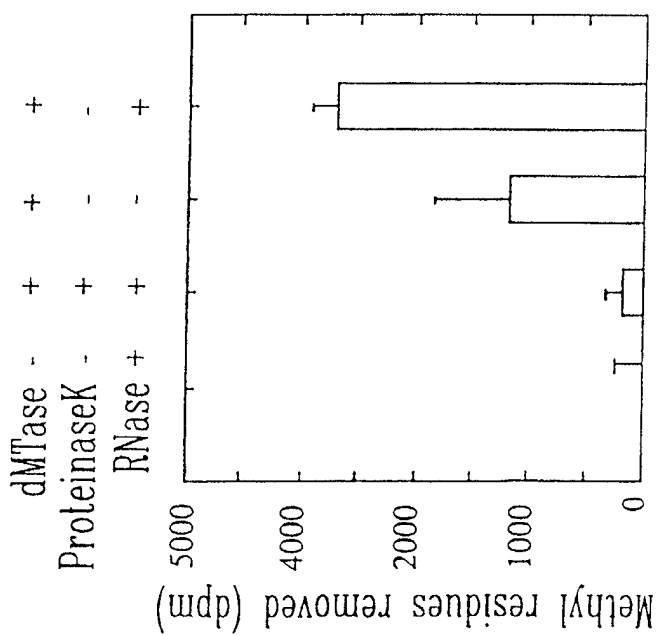


FIG. 21

5/50

FIG. 3A

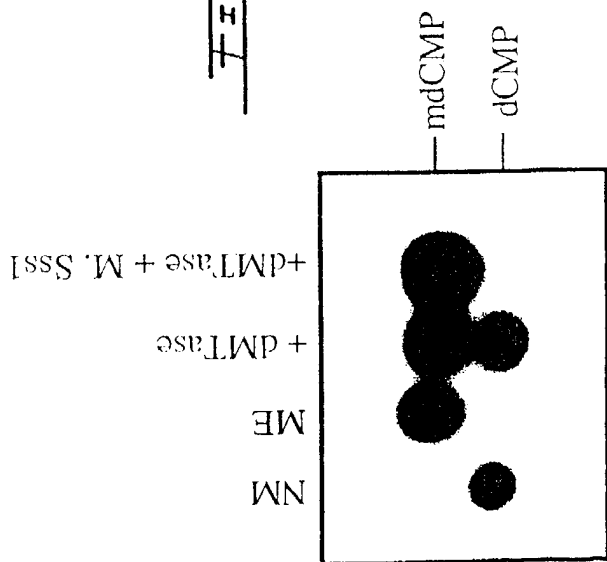
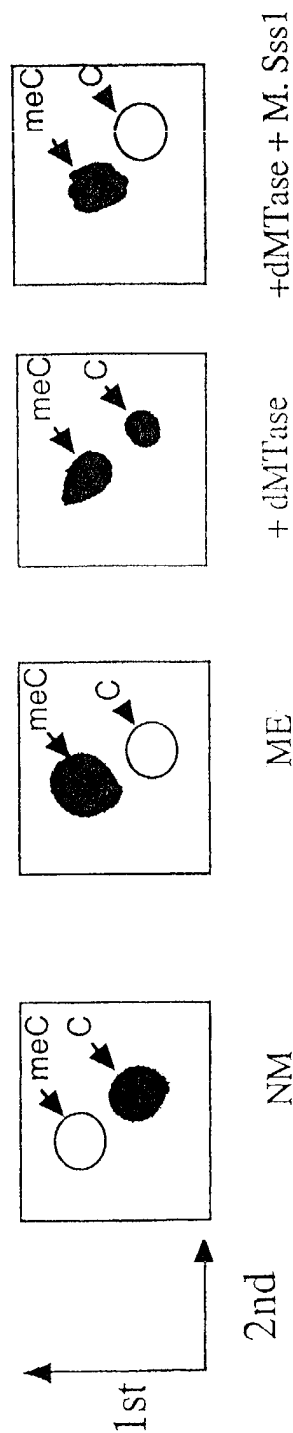
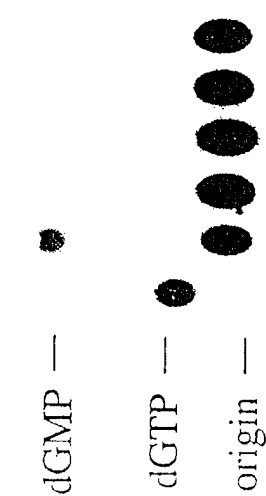


FIG. 3B



6/50

- V PDS
[N.E. 0 60 120 +RNase]
dGTP



Labeled nucleotide:
[α^{32} P]-dGTP

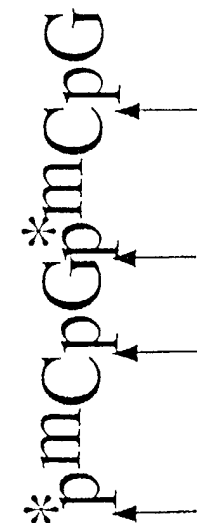


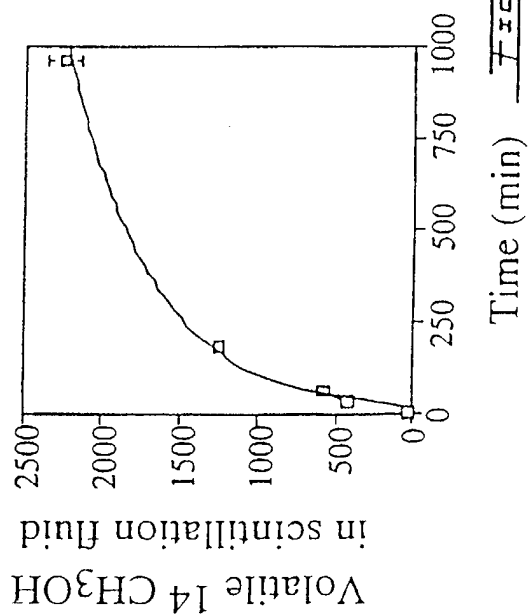
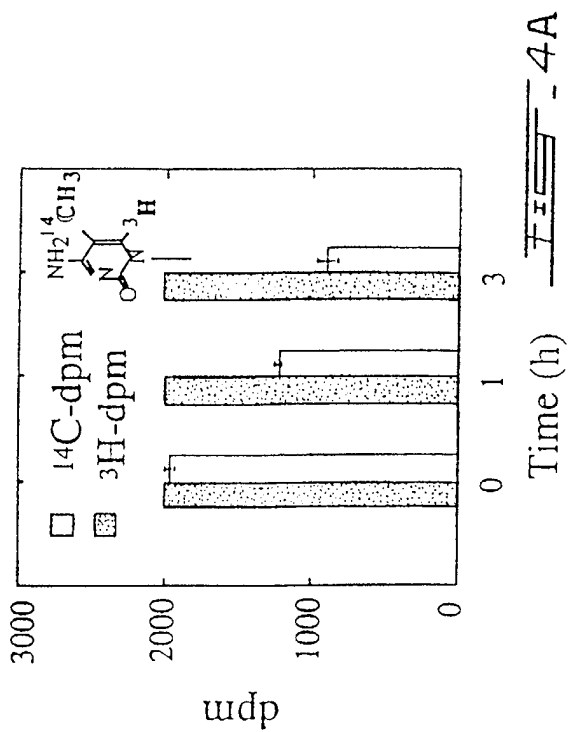
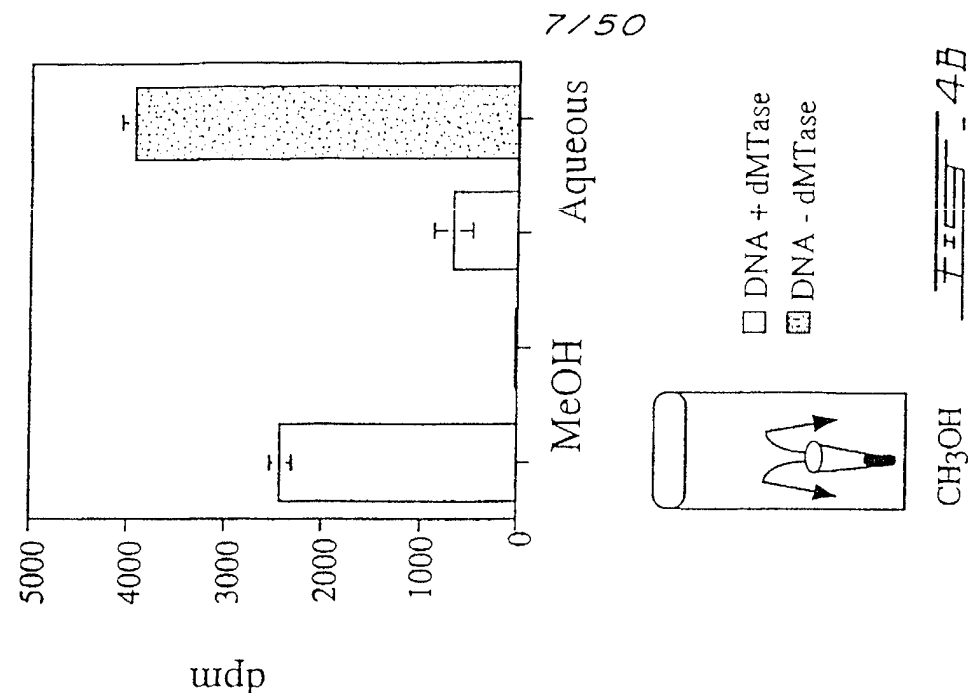
7E-3C

+ V PDS
[+RNase -RNase NM MB]
mdCMP —
dCMP —

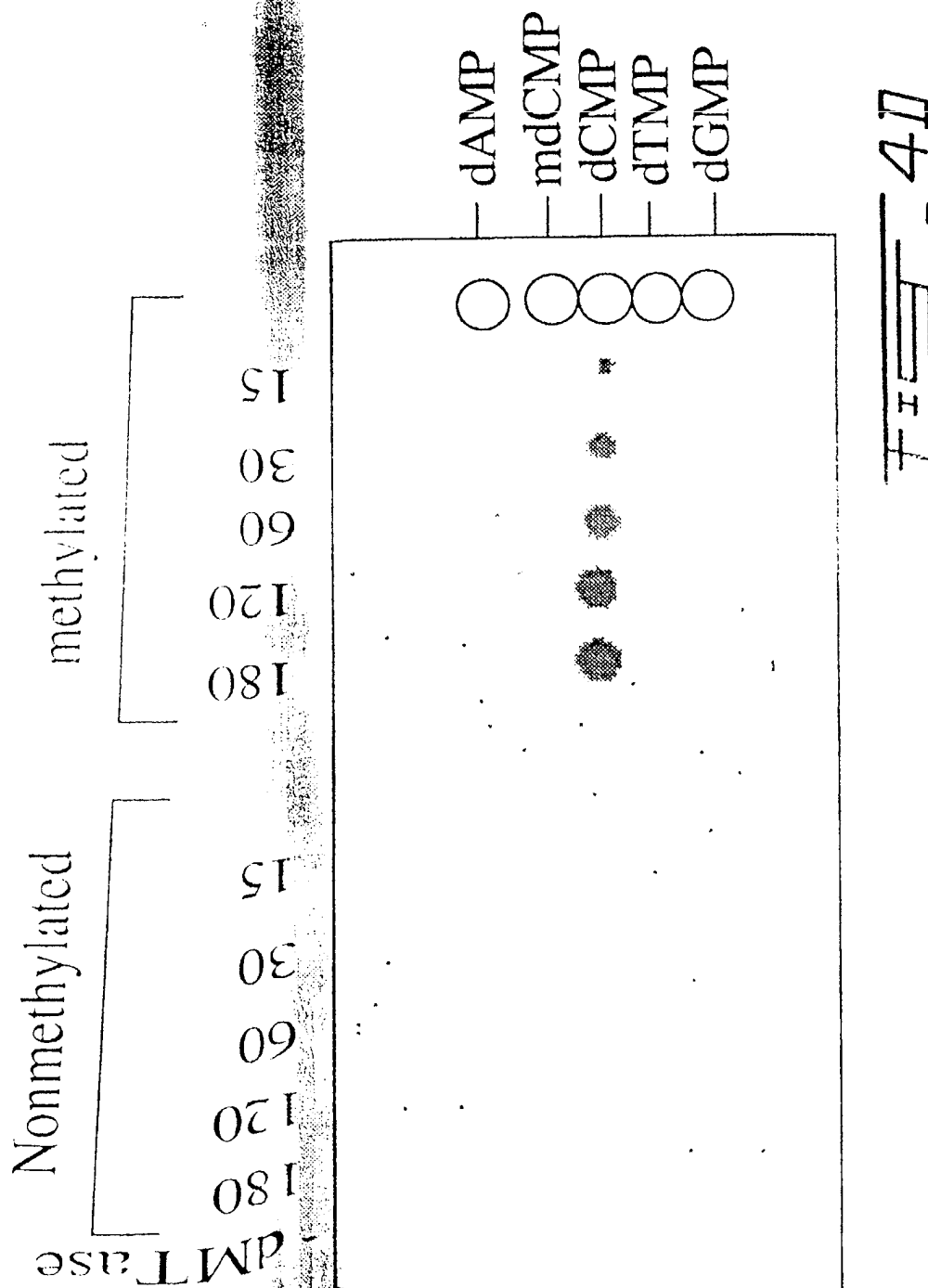


Labeled nucleotide:
[α^{32} P]-dCTP

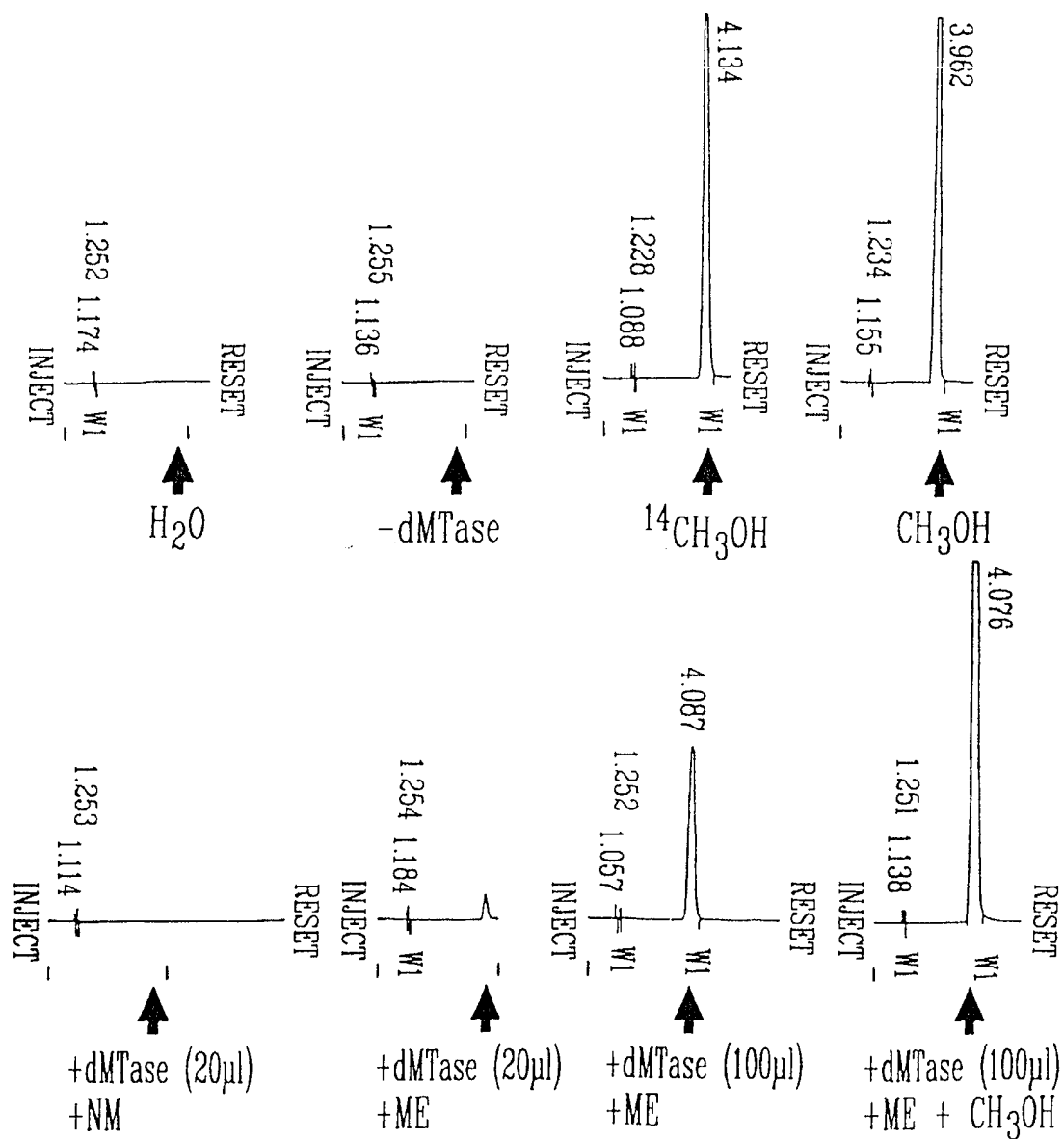




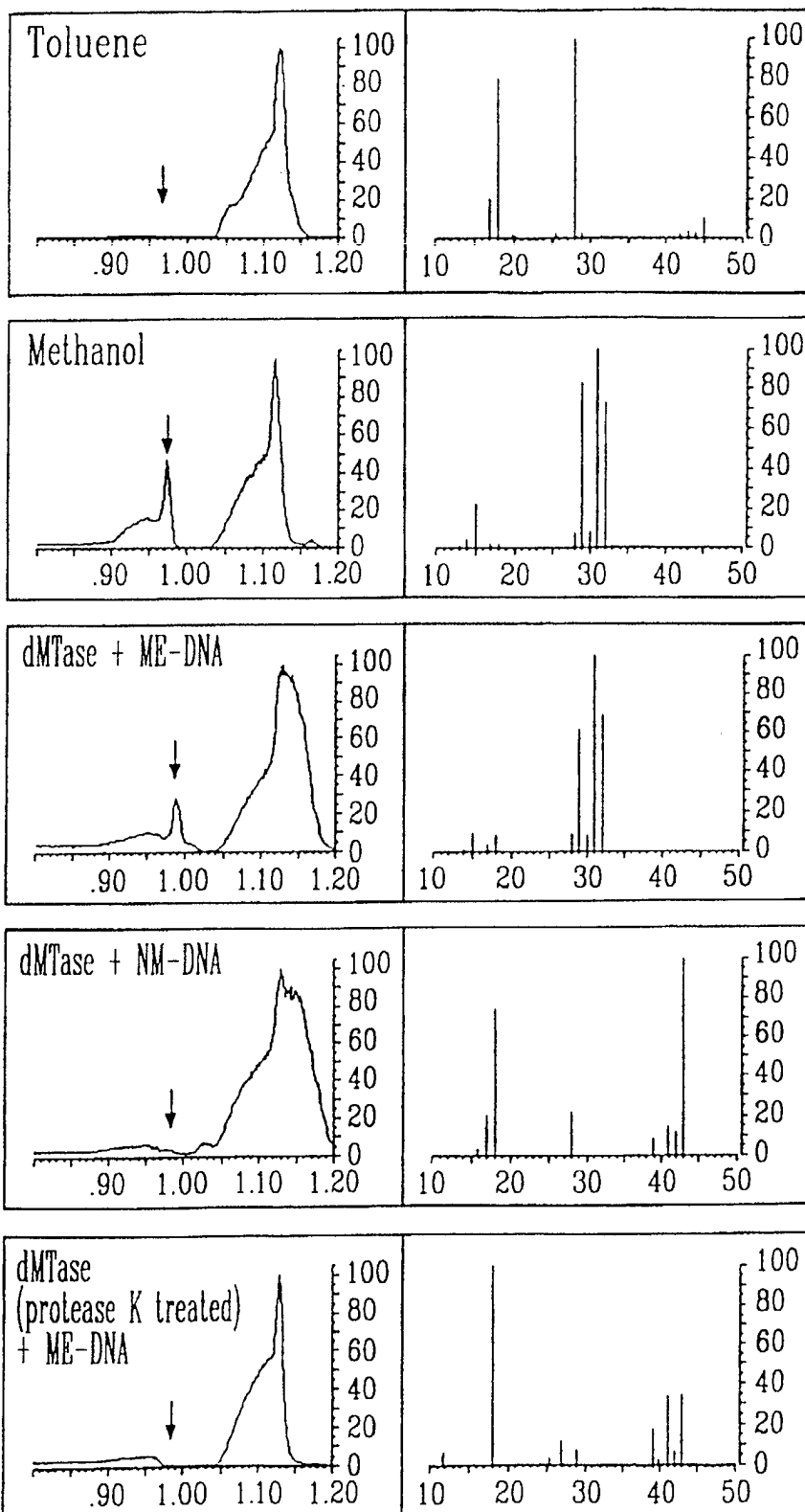
8/50



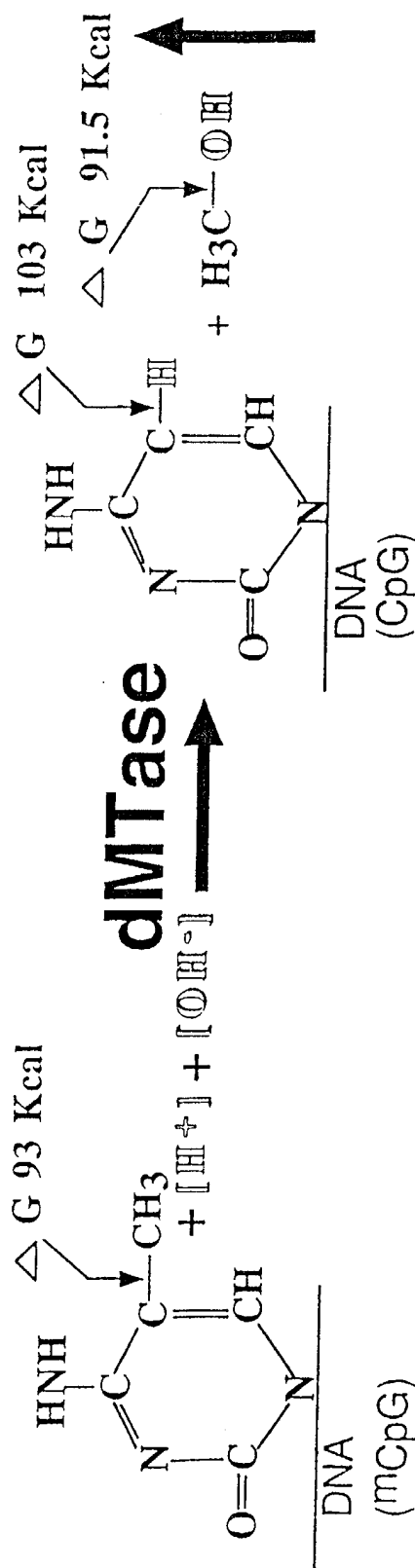
9/50

FIG. 4E

10/50

FIG. 4F

11/50



$$\Delta G_{\text{Reaction}} = (93) - (103 + 91.5) = (-101.5) \text{ Kcal}$$

FIG. 5

12/50

| M.MspI+M.SssI | | dMTase | | | | nondigested plasmid | |
|---------------|---|--------|---|---|---|------------------------|--|
| - | + | - | - | + | - | + | |
| EcoRII | + | + | + | + | + | + | |
| DpnI | + | + | + | + | + | + | |
| HpaII | + | + | + | + | + | + | |
| HhaI | + | + | + | + | + | + | |
| MspI | + | + | + | + | + | + | |

0.45 0.57
0.40 >
0.39 0.27
0.34 0.10
0.33

1.00
0.71
0.49
0.32
0.24
0.15

13/50

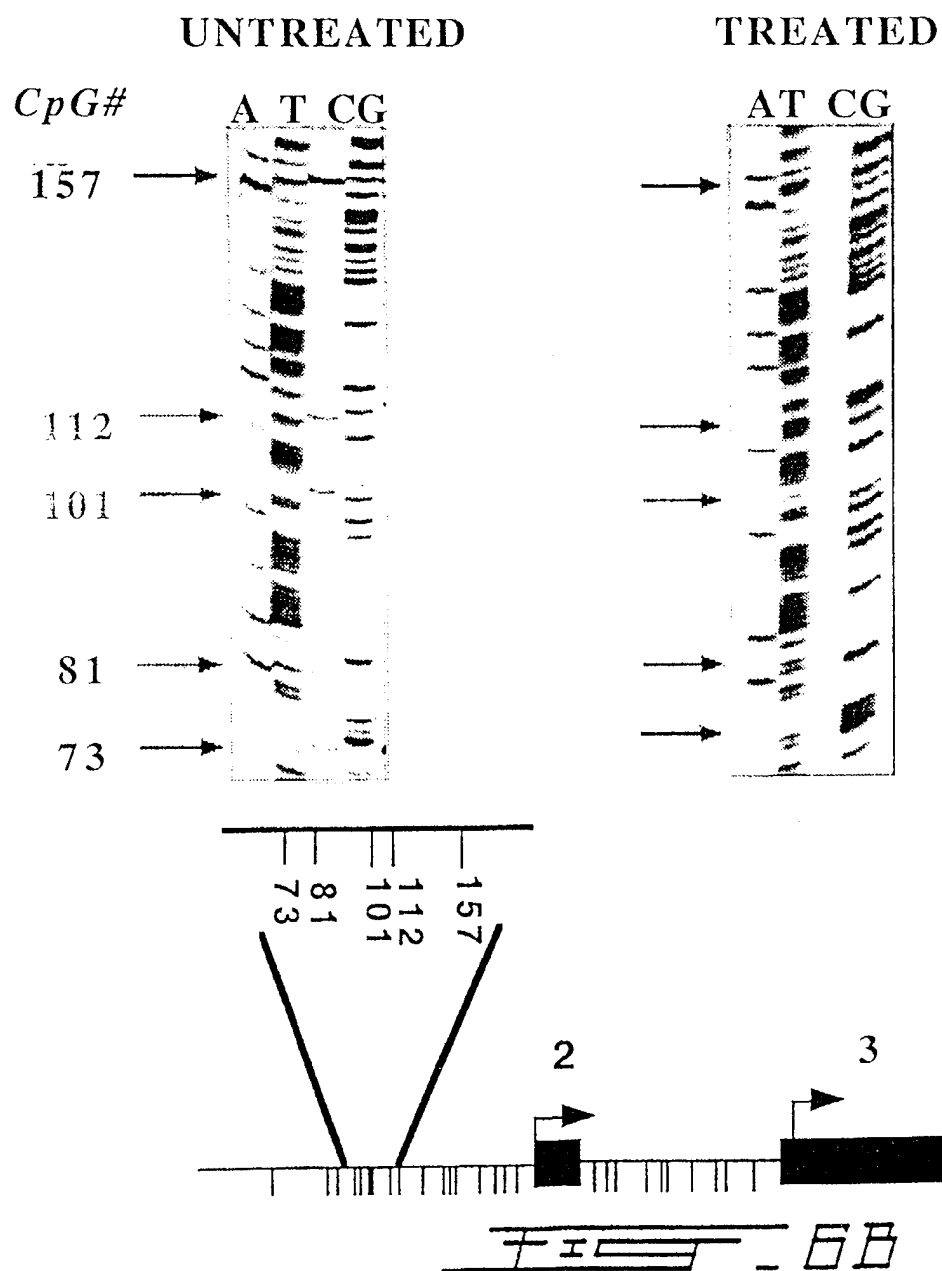
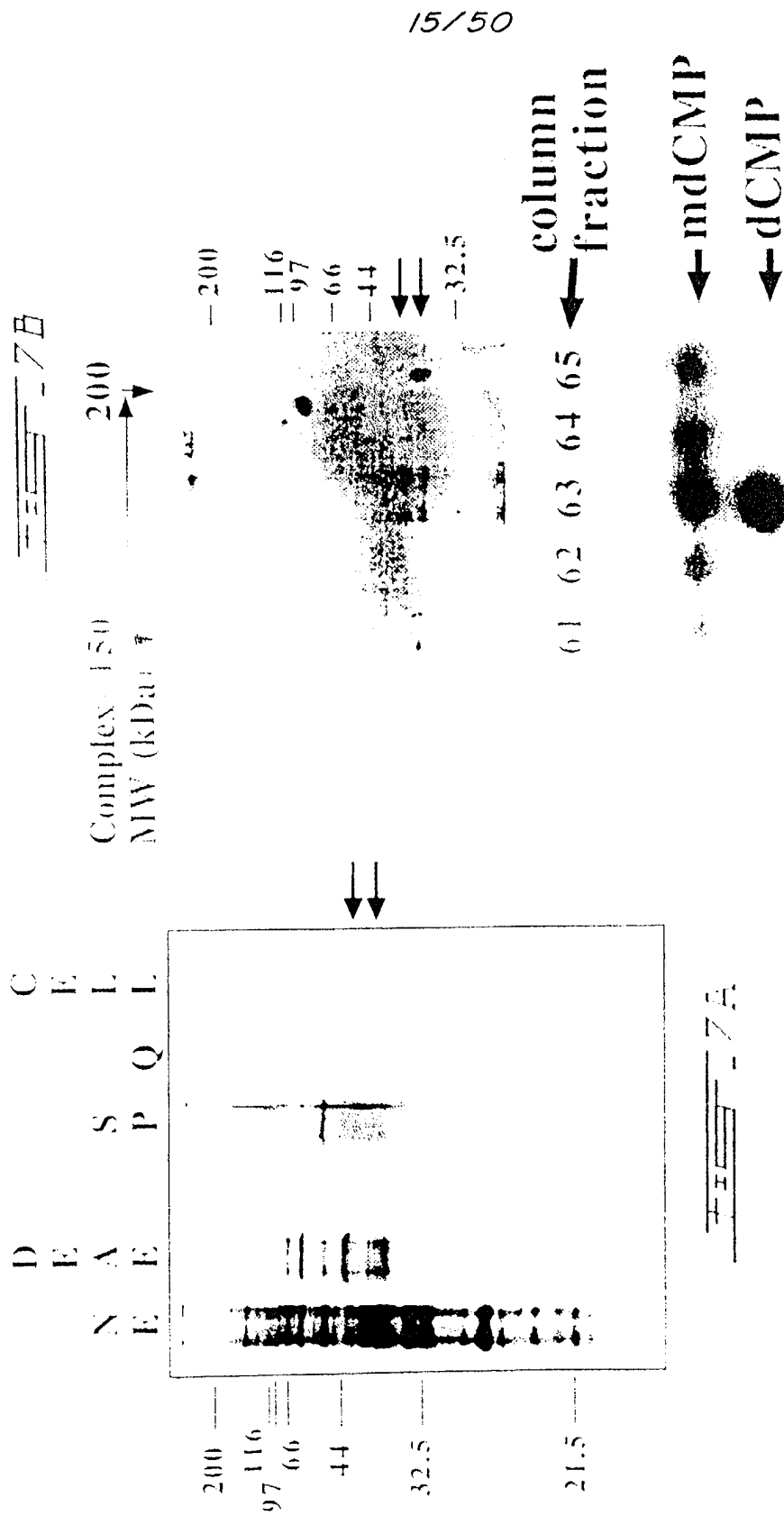


Figure 1 shows a series of autoradiographs of DNA bands over time (0, 0.5, 1, 2, 3, 18 hours) for four different sequences: A, meC, C, and T. The bands are labeled with their respective nucleotides and arrows. The time points are indicated at the bottom. The sequence A-T-C-G is shown at the bottom right.



16/50

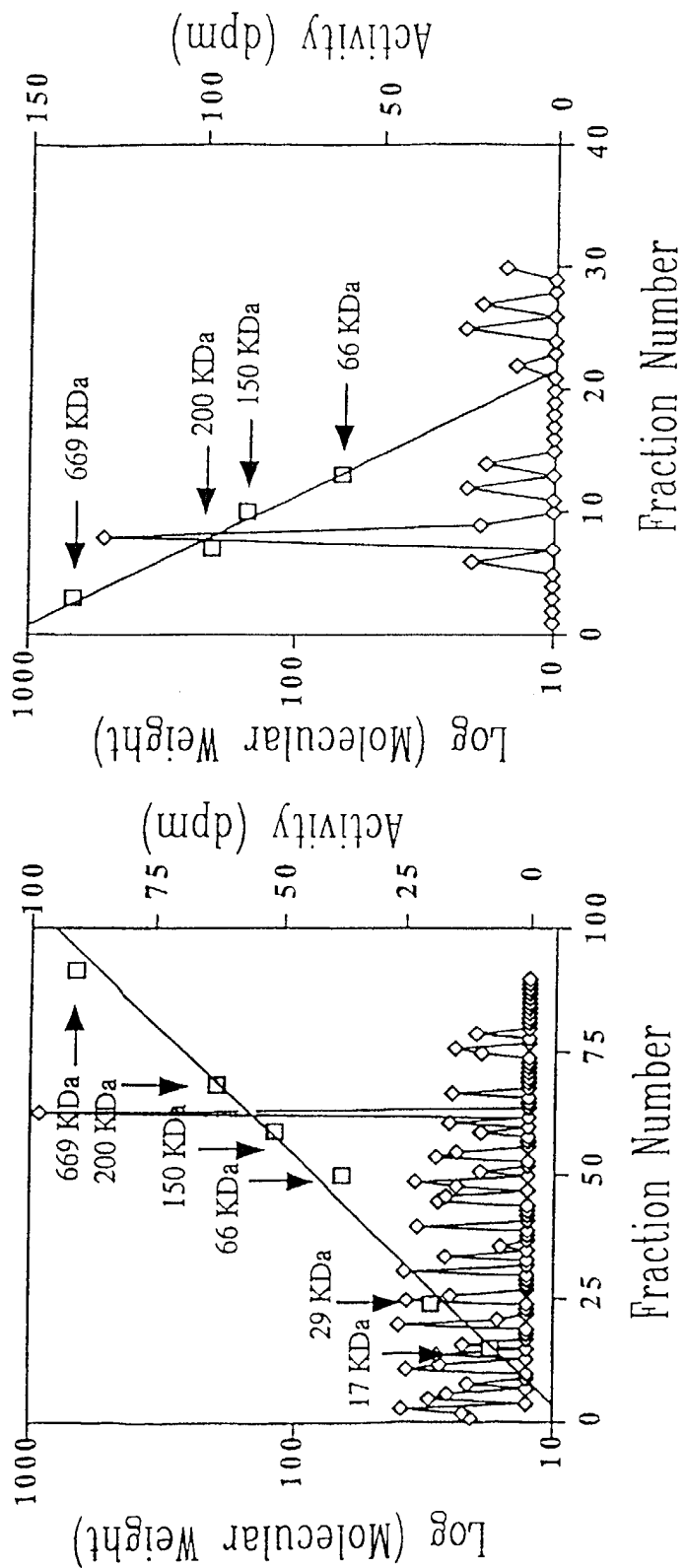


Fig. 7D

Fig. 7C

17/50

EST DCPALPPGWKKEEVIRKSGLSAGKSDVYVFSPSGKKFRSKPQLARYLGNIVDLS
 |||||
 D P LP GW ++ RKSG SAGK DVY +P GK FRSK +L Y D S
 |||||
 MeCP2 15 DDPTLPEGWIRKIKQKRSGRSAGKYDVYLINEPQKAFRSKVVELIAYFEKVGDTS 68

FEF - BA

| | | |
|---|-----|--|
| MDCPALPPGW KKEEVIRKSG LSAGKSDVY FSPSGKKFRS | 40 | |
| | | |
| KPOLARYLGN TVDLSSFFDR TGKMMPSKLQ KNKQRLRNNDP | 80 | |
| | | |
| LNQNKGGKPD L NITLPIRQTA SIFKQPVTKV TNHPSNKVKS | 120 | homology to methylated DNA binding domain |
| DPQRMNEQPR QLFWEKRLQG LSASDVTEQI IKTMELPKGL | 160 | |
| QGVGPGSND E TLLSAVASAL HTSSAPITGQ VSAAVEKNPA | 200 | homology to coiled coil domain |
| VWLNTSQPLC KAFIVTDEDI RKQOEERVOOV RKILEDALMA | 240 | |
| | | |
| DILSRAADTE EMDIEMDSGD EA | 262 | |

FEF - BB

18/50

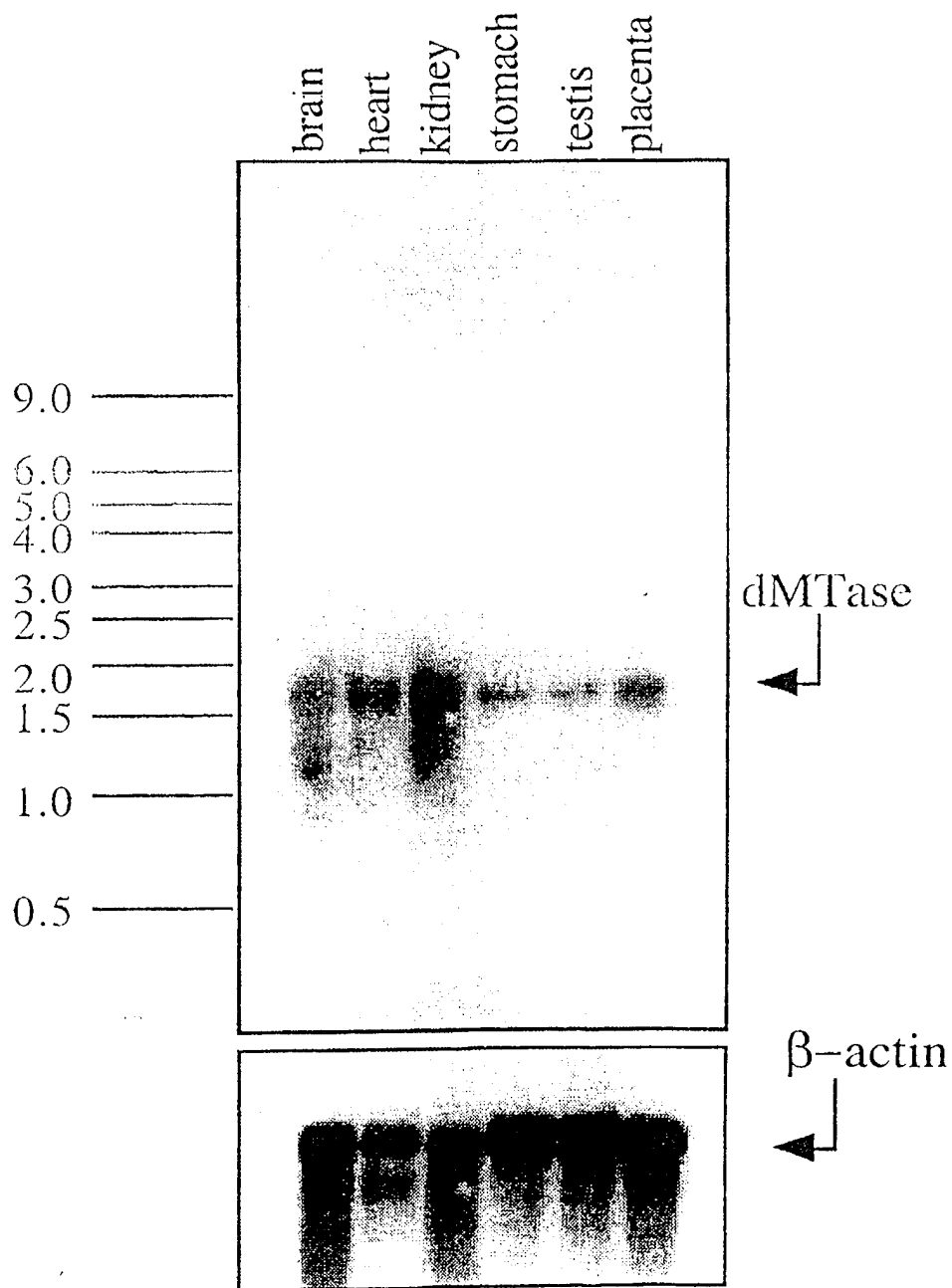


FIG. 8C

Human DNA demethylase cDNA-dMTase1 and predicted amino acid sequence

5' gggggcggtgg ccccgagaag gcggagacaa gatggccgcc catagcgctt ggaggaccta
 agaggcggtg gccggggcca gcccccggc aggagggccg ctctgtgcgc gcccgctcta
 tgatgcttgc gcgcgtccc gcgcgccgc agcgcacgc cggctggga gggcggtct cccggattcc
 aagggtcgg ttacggaaga ggcgactgct gcccgagca ggagggctgga tgcgcgcgca
 cccggggga gcccgctgct ggcgactccg ccatagaca gggggccag gagagtgcgg cggcggcag
 cggcgctggc ggcgtgcga ggcgtgcga ggcgtggc gccagtggc tgcgccgtc
 cccggtgagc ggcggggcg gccggggcg gcggcgtctg tggccgtggc gggggcggtg
 gaagcaggcg gccggggcg gccggggcg gccggggcg cggggccggg gccgtggcg
 gggacgggga cggggccggg gccggggcg gccggggcg cggcggtggc gcagcggcct
 tggcggcgac ggcggcggc cctttccgt gccagcggc gccggcggc ccccccggc
 ggagccggtc aagaggatgg attgccggc gccggccggg ccagggggac cccggggccac
 ggagagcggg gatccgaaa tctgggctaa gtgctggcaa gacgatgtc tactactca gtccaagtgg
 taagaagtgc agaagcaagc ctcagttggc aaggtacctg ggaatactg ttgatctcag
 cagttttgac ttcagaactg gaaagatgat gcctagtaa ttacagaaga acaaacagag
 actgcgaaac gatcctctca atcaaaataa gggtaacca gacttgaata caacattgcc
 aattagacaa acagcatcaa ttttcaaca accggtaacc aaagtcaaca atcatcctag

19/50

FEF-9A

20/50

taataaagtg aaatcagacc cacaacgaat gaatgaacag ccacgtcagc ttttctggga
gaagaggcta caaggactta gtgcatacaga tgtaacagaa caaattataa aaaccatgga
actacccaaa ggtcttcaag gagtgggtcc aggtagcaat gatgagaccc ttttatctgc
tgttgccagt gcttgcaca ccaagctctgc gccaatcaca gggcaagtct ccgctgctgt
ggaaaagaac cctgctgtt ggcttaacac atctcaaccc ctctgcaaaag cttttattgt
cacagatgaa gacatcagga aacaggaaga gcgagtacag caagtacgca agaaattgga
agaagcactg atggcagaca tcttgtcgcg agctgctgat acagaagaga tggatatgga
aatggacagt ggagatgaag cctaagaata tgatcaggta acttcgacc gactttcccc
aagrgaaaat tcctagaaat tgaacaaaaa tgtttccact ggcttttgcc tgtaagaaaa
aaaatgtacc cgagcacata gagcttttta atagcactaa ccaatgcctt tttagatgta
tttttgatgt atatatctat tattcaaaaa atcatgttta ttttgagtcc taggacttaa
aattagtctt ttgtaatatc aagcaggacc ctaagatgaa gctgagcctt tgatgccagg
tgcaatctac tggaaatgta gcacttacgt aaacatttg tttccccac agtttaata
agaacagatc aggaattcta aataaatttc ccagttaaag attattgtga cttcactgta
tataaacata tttttatact ttattgaaag gggacacctg tacattcttc catcatcact
gtaaagacaa ataaatgatt atattcacia aaaaaaaaaa 3'

SEQ ID NO:1

FIG. 1 - 9B

21/50

MRAHPGGGCCPEEGESAAAGSGAGGDSAIEQGGQSALAPSPVSGVR
REGARGGGRGRWKQAGRGGVCGRGRGRGRGRGRGRGRGRPPSG
GSGLGDDGGGGGGGAPRRPVPFPSSAGPGRGPRAATESGKRM
DCPALPPGWKKEEVIRKSGLSAGKSDVYFSPSGKKFRSKPQLARYLGNT
VDLSSDFRTGKMMPSKLQKNKQRLRNDPLNQNKGKPDNLN'TLPIRQTAS
IFKQPVTKVTNHPSNVKSDPQRMNEQPRQLFWEKRLQGLSASDVTEQII
KTMELPKGLQGVGPGSNDETLLSAVASALHTSSAPI TGQVSAAVEKNPAV
WLNTSQPLCKAFIVTDEDIRKQEERVQQVRKKLEALMADILSRAADTEE
MDIEMDSGDEA

SEQ ID NO: 2

FEI - 9C

Human DNA demethylase homologue-dMTase2 and predicted amino acid sequence

5' agcgggccga ggagccgggc gcaatggagc ggaagaggtg ggagtgcccg gcgtcccg
 agggctggga gaggaagaa gtgcccagaa ggtcggggct gtggccggc cacagggatg
 tcttttacta tagcccgagc ggaagaagt tccgcagcaa gccgagctg gcgcgtacc
 tgggcggctc catggacctg agcaccttcg acttcgcac gggcaagatg ctgatgagca
 agatgaacaa gagcggccag cgcgtgcgct acgactcctc caaccaggtc aaggcaagc
 ccgacctgaa cacggcgctg cccgtgcgcc agacggcgtc catctcaag cagcgggtga
 ccaagattac caaccacccc agcaacaagg tcaagagcga ccgcagaag gcggtggacc
 agccgcgcca gctctcttgg gagaagaagc tgagcggcct gaacgccttc gacattgctg
 aggagctggt caagaccatg gacctccca agggcctgca gggggtggga cctggctgca
 cggatgagac gctgctgtcg gccatgcga gcgcctgca cactagcacc atgcccatca
 cgggacagct ctggccgccc gtggagaaga acccggcgt atggctcaac accacgcagc
 ccctgtgcaa agccttcatg gtgaccgacg aggacatcag gaagcaggaa gagtgggtgc
 agcaggtgctg gaagcggctg gaggaggcgc tgatggccga catgtggcg cacgtggagg
 agctggcccc tgacggggag gcgccgtgg acaaggcctg cgctgaggac gacgacgagg
 aagacgagga ggaggaggag gaggagcccc tgctggagcc gcctgcagac gcggtcctcg
 gcagaggccc tgccgagagc ccgtgctgcc ccagccttg gagacacca ggaggaaggc
 gccccacgtg aaccaggctc ggccggcgaag ccacttccc gggcctcggg gcacacagct
 cgtgctcctg gctccctcct cggcccgtcc ccacttccc gctcgctcct tacagagtcc ggcttcggga
 ggggctgccc ccaccgaaa gacctccac gctcgctcct cctctacgac ctttgggctc gaggccagct
 agtgcggggt gctcctgggc cctgcctggc tccctacgac ctttgggctc gaggccagct

22/50
 715-90

23/50

cctccccatg cccgctgtcc cagctccttg agactggaga gcagccagca ggtgccccggc
agctcggcgc cacggcttgc tgacagctgg gaggtttctt cggctctggag gcgtagtcttt
gaaactcaca tcaccactg tcacccactg ggcagcgtga ggacgggact ctggtctgct gtgggggggca
tgcaggacgg cgccactctc cgccactctc tgccctgcca tgcggctggt ggtgccacag agcctcacgg
tgccctgagtg gcgtgcccag ggaggccgct ctccctcagt aaatgtaaca cagtcgaggc
acgtcatcgg gcagcccttcc ctgtgtgcca acgccagcct tcgcttctga aaacaaact
ccagcccgctg ccagtccgga ccttggtcgcc cggcgctgcc agaattgctcc actgccagcc
ggccccccctg cctcggtttc ccttctgttt agtggcgaca caggcaccca gctttgggggt
ggtgctgacg ctccaggggg tgccaggagc cactgggaca gggtagggct ccagacgct
cctcgaggtg ccagctctc caggagctt ctggcccaag gcgttcttga gggatctgct
ccttaacccc ccagtgcctt ggcgagggca ggttccaagc cacagacgcc tgccccgagt
ggactttgcg gccagtccct gggcgcttc ggtggccctg cttgcccagt gagggttccct
aacgggtggg ttcawtggcc tggcccvagc gagccccac ctgcattgac cttaggccca
tagagagggc ctgtcccggc gctgcccag ccaaggatct ggtcgctgcc ccagggggac
tgatggggcaa gagtgcgcc tgtggctgga ctgtgaccat ccctgatggg gctgaccgc
gggagctgag gaagcgccgc tcaccgtct gccctccaag gaccgcatg gaggcagtgg
gctggcagct tcctgctgct ccctgtcaga gtcaaaagcac aaatcctcag gacgggctca
agggccaggg cagccgaggg aagctccagg tggggaccac gtcttctga ggttggtgcc
cactggctgg gaccctttgc agtggggtgg cctccccctt gtctgcctgg tggagggagc
cgtgggcgtg gggacgtgac tgaataaagc caccatgggt ggatgtgctt gg 3'

SEQ ID NO:3

FEF-9E

24/50

MERKRWECPALPQGWEREVPRRSGLSAGHRDVFYSPSGKKFRSKPQLA
RYLGGMDLSTFDFTGKMLMSKMNKSQRQVRDYDSSNQVKGKPDNLNTALP
VRQTASIFKQPVTKITNHPSNVKSDPQKAVDQPRQLFWEKKLSGLNAFD
IAEELVKTMDLPKGLQGVGPGCTDETLLSAIASALHTSTMPITGQLSAAV
EKNPGVWLNTTQPLCKAFMVTDEDIRKQEEELVQQVRKRLEEALMADMLAH
VEELARDGEAPLDKACAEDDDDEEEEEEPDPPEMEHV

SEQ ID NO: 4

FIG. 1

Lipman-Pearson Protein Alignment

Ktuple: 2; Gap Penalty: 4; Gap Length Penalty: 12

| Seq1(1>411) | | Seq2(1>291) | | Similarity | | Gap | | Consensus | |
|--|-----------------------|-----------------------|-----------------------|------------|--------|--------|--------|-----------|--------|
| human dMTase1 protein | human dMTase2 protein | human dMTase2 protein | human dMTase2 protein | Index | Number | Length | Length | Length | Length |
| (148>397) | (4>253) | (4>253) | (4>253) | 76.0 | 0 | 0 | 0 | 250 | 250 |
| (148>397) | (4>253) | (4>253) | (4>253) | 76.0 | 0 | 0 | 0 | 250 | 250 |
| v150 | v160 | v170 | v180 | v190 | v200 | v210 | | | |
| KRMDCPALPPGWNKKEEVIRKSGLSAGKSDVYFSPSGKKFRSKPQLARYLGNIVDLSSDFRTGKMMPSK | | | | | | | | | |
| KR : CPALP.GW.:EEV R:SGLSAG...DV:Y:SPSGKKFRSKPQLARYLG.:DLS:FDFRTGKM:SK | | | | | | | | | |
| KRWECPALPQGWEREIEVPRRSGLSAGHRDVFYSPSGKKFRSKPQLARYLGGMDLSIFDFRTGKMIMSK | | | | | | | | | |
| ^10 | ^20 | ^30 | ^40 | ^50 | ^60 | ^70 | | | |
| v220 | v230 | v240 | v250 | v260 | v270 | v280 | | | |
| LQKNKQRLFNNDPLNQNKGPDLNNTLPTRQTASIFKQPVTKVINHPNKNKVSDFQRMNEQPRQLFWFKRL | | | | | | | | | |
| ::K::QR:R D: NQ KGKPDINT:LP:RQTASIFKQPVTK:TNHPNKNKVSDFQ: :QPRQLFWFK:L | | | | | | | | | |
| MNKSQRVRVYDSSNQKGKPDINTALPVRQTASIFKQPVTKITNHPNKNKVSDFQKAVDQPRQLFWFKKL | | | | | | | | | |
| ^80 | ^90 | ^100 | ^110 | ^120 | ^130 | ^140 | | | |

 96

v290 v300 v310 v320 v330 v340 v350
 QGLSASDVTEQI IKIMELPKGLQGVGPGSND EILLSAVASALHTSSAPITGQVSAAVEKNP VAWINITSQP
 GL:A D::E::KIM:L PKGLQGVGPG..DETLISA:ASALHTS: PITGQ:SAAVEKNP:VWINT:QP
 SGLNAFDIAEELVKIMDL PKGLQGVGPGCTDETLISAISALHTSIMPITGQLSAAVEKNP VAWINTTQP
 ^150 ^160 ^170 ^180 ^190 ^200 ^210

26/50

v360 v370 v380 v390
 LCKAFIVTDEDIRKQEERVQQVRKKLEEFALMADILSRAAD
 LCKAF:VTDEDIRKQEE VQQVRK:LEEFALMAD:L:::
 LCKAFMVTDEDIRKQEELVQQVRKKLEEFALMADMLAHVEE
 ^220 ^230 ^240 ^250

FILE - 9H

Mouse DNA demethylase-dMTase1 and predicted amino acid sequence

27/50

```

5' ccgctctgcg ggcgggcgcg gtctccggga ttccaagggc tcggttacgg aagaagcgca
gagccggctg gggagggggc tggatcgcg cgcaccggg gggaggccgc tgctgcccgg
agcaggagga ggggagagag cgcgggggcg gcagcggcgc tggcggcgac tccgccatag
agcagggggg ccagggcagc gcgctcgctc cgtccccggt gagcggcggtg cgcagggaag
gcgctcgggg cgcgggcggt ggcggggggc ggtggaagca ggcggcccg ggcggcgcg
tctgtggcgg tggcgtggc cgtggccggg gtccggggcg tggcggggc tggcggcgcg
gccgcgcccg tcccagagt gccggcagc ggtggcgcg cccttggcg cgacggcgcg ggcggcgcg
gcggctgcgg cgtcggcagc ggtggcgcg tggggcgcg tgcggcgcg gcgggacct gtcccttcc
cgtcggggag ctcgggggcg gggcccagg gaccccggg caccgagagc ggaagagga
tggactgccc ggcctcccc cccggatgga agaaggagga agtgatccga aaatcagggc
tcagtgtctg caagagcgat gtctactact tcagtccaag tggtaagaag ttcagaagta
aacctcagct ggcaagatag ctgggaaatg ctgttgacct tagcagtttt gacttcagga
ccgggcaagat gatgcctagt aaattacaga agaacaagca gagactccgg aatgaccccc
tcaatcagaa caagggtaaa ccagacctga acacaacatt gccaataga caaactgcat
caattttcaa gcaaccagta accaaattca cgaaccacc gagcaataag gtgaagtcag
accccagcg gatgaatgaa caaccacgtc agcttttctg ggagaagagg ctacaaggac
ttagcgcatac agatgtaaca gaacaaatta taaaaccat ggagctacct aaaggtcttc
aaggagtcgg tccaggtagc aatgacgaga cccttctgtc tgctgtggcc agtgctttac

```

FEI-91

28/50

acacaagctc tgcgcccatc acaggacaag tctctgctgc cgtggaaaag aaccctgctg
tttggcttaa cacatctcaa cccctctgca aagctttcat tgttacagat gaagacatta
ggaacacagga agagcgagtc caacaagtac gcaagaaact ggaggaggca ctgatggccg
acatcctgtc ccgggctgcg gacacggagg aagtagacat tgacatggac agtggagatg
aggcgtaaga atatgatcag gtaactttcg actgacctc cccaagagca aattgctaga
aacagaatta aacatttcc actgggttcc gcctgtaaga aaaagtgtac ctgagcacat
agctttttaa tagcactaac caatgccttt ttagatgtat ttttgatgta tatactatt
attcccaaatg atgtttattt tgaatcctag gacttaaaat gagtctttta taatagcaag
cagggccctt ccggtgcagt gcagcttga ggccaggctgc agtctactgg aaaggtagca
cttacgtgaa atatttgttt cccccacagt tttaatatata acagatcagg agtaccaat
aagtttccca attaaagatt attatacttc actgtatatata aacagatttt tatactttat
tgaaagaaga tacctgtaca ttcttccatc atcactgtaa agacaaataa atgactatat
tcac 3'

SEQ ID NO:5

7-5-91

29/50

MRAHPGGRCCEEEGESAAGSGAGGDSAIEQGGQGSALAPSPVSGVR
REGARGGRGRWKQAARGGVCGRGRGRGRGRGRGRGRGRGPQSG
GSLGGDGGGAGGCGVGGGVA PRRDVPFPSSGSGPGRPRATESG
KRMDCPALPPGWKKEEVIRKSLGAGKSDVYYFSPSGKKFRSKPQLARYL
GNAVDLSSFDFTGKMMP SKLQKNKQRLRNDPLNQNKGKPD LNTTLPIRQ
TASIFKQPVTKFTNHPSNKVKSDPQRMNEQPRQLFWEKRLQGLSASDVTE
QIIKT MELPKGLQGVGPGSND ETL LSAVASALHTSSAPITGQVSAAVEKN
PAVWLNTSQPLCKAFIVTDEDIRKQEEERVQQVRKKLEEALMADILSRAAD
TEEV D IDMDSGDEA

SEQ ID NO: 6



30/50

Mouse DNA demethylase-dMTase2 and predicted amino acid sequence

```

5' cacgcgcggg cgggtgggcg gagcgcccc ctagcgggg gctgtgaagc gcggggagg
ggcgcgcgcg gtggcgagc cggcgcgcg cggctgggg gcggaggcg gaggccctg
ggacagaaca gctgcggcga gtggcggcg cggaggagc cgaatcggcg acgagcccg
gggtcgcaac ttgcagaagc ggcggcgcg cgggcaccg ccacggcgg cggaagacc
ggggcgcaat ggagcggaag aggtgggagt gcccgcgct ccgcaggc tgggaagg
aagaagtgcc caggaggtcg gggctgtcg ccggccacg ggatgtctt tactatagc
ccagcgggaa gaagtccgc agcaagccac aactggcac ttacctggc ggatccatg
acctcagcac ctgcacttc cgcaccgaa agatgttgat gaacaagatg aataagagtc
gccagcgtgt gcgctatgat tctccaacc aggtcaagg caagcctgac ctgaacacc
cgctgcctgt acggcagact gcaccatct tcaagcaacc ggtgaccaag atcaccaac
acccagcaa caaggtcaag agcagcccg agaaggcagt ggaccagcg aggcagctt
tctgggagaa gaagctaagt ggattgagt cctttgacat tgcagaagaa ctggtcagga
ccatggactt gcccaaggc ctgcaggag ctggccctgg tggacagat gagcgtgc
tgtcagccat tgcgagtgc ctacacacca gcaccctgcc cattacaggc cagctctctg
cagccgtgga gaagaaccct ggtgtgtggc tgaacactgc acagccactg tgcaagcct
tcatggtgac agatgacgac atcaggaagc aggaggagct ggtacagcag gtacggaagc
gcctggagga ggcactgatg gccgacatgc tagctcatgt ggaggagct gccgagacg
gggaggcacc actggacaag gcctgtgcag aggaggaaga ggaggaggaa gaggaggagg

```

755 - 91

31/50

aagagccgga gccagagcga gtgtagcaca ggtgccctgc ccaagtctgg gctgcagact
gccttcagcc ttgcctggac caggtagggg ccagacctgt aggaggcagc cgtccacctc
ctttccaaag cctcctgctt ccaggtctca gtgcaggagg cccctgtgga ccttgaactc
acttgtccct gcgctgcctg gcaggaagcc ccacactgaa agcagatgag cagtgaacca
actgagaggg cacctggaca cagtcacctc cctgcctcct tatcatagga caaggccttg
cttggcaccg aggagctggg agccgtgttg ggtgctggag gaagtctctg gaaacacacc
tggctatgcc caccttatgt ccctaaggct attacaggcc agggtttgga ctgctccggc
ccacagggct gccagcctc cccacactga ggttcagcag ccaccaggga agtcactttc
cttcaataaa ctgatggtag gaacttgtg 3'

SEQ ID NO:7

~~FILE~~ - 9M

32/50

MERKRWECPALPQGWEEEEVPRRSGLSAGHRDVFYSPSGKKFRSKPQLA
RYLGGSMDLSTFDFTGKMLMNKMNKSRQVRDYDSSNQVKGPDLNTALP
VRQTASIFKQPVTKITNHPSNVKSDPQKAVDQPRQLFWEKKLSGLSAFD
IAEELVRTMDLPKGLQGVGPGCTDETLLSAIASALHTSTLPITGQLSAAV
EKNPGVWLNTAQPLCKAFMVTTDDDIRKQEELVQQVRKRLEEALMADMLAH
VEELARDGEAPLDKACAEEEEEEEEEEEEEPEPERV

SEQ ID NO: 8

7-15-59 NG

Consensus
Length

Length

250 0

3

DATA

0/0

1

THE

0471

1

11

34/50

v300 v310 v320 v330 v340 v350 v360
 QGLSASDVTEQI IKIMELPKGLQGVGPGSND EITLSA VASALHTSSAPITIGQVSA AVEKNP A VWLNTS QP
 GLSA D::E:::TM:L PKGLQGVGPG..DEITLSA:ASALHTS: PTTGQ:SA AVEKNP: VWLNT:QP
 SGLSAFDIAEELVRIMDL PKGLQGVGPGCITDEITLSA IASALHTSITLPITIGQLSA AVEKNP G VWLNTA QP
 ^150 ^160 ^170 ^180 ^190 ^200 ^210

v370 v380 v390 v400
 LCKAFTVDEDIRKQEEERVQ QVRKKLEEFALMADILSRAAD
 LCKAF:VTD:DIRKQEE VQ QVRK:LEEFALMAD:L::: :
 LCKAFMVTDDDIRKQEEELVQ QVRKKLEEFALMADMLAHVEE
 ^220 ^230 ^240 ^250

T E E - 9 P

35/50

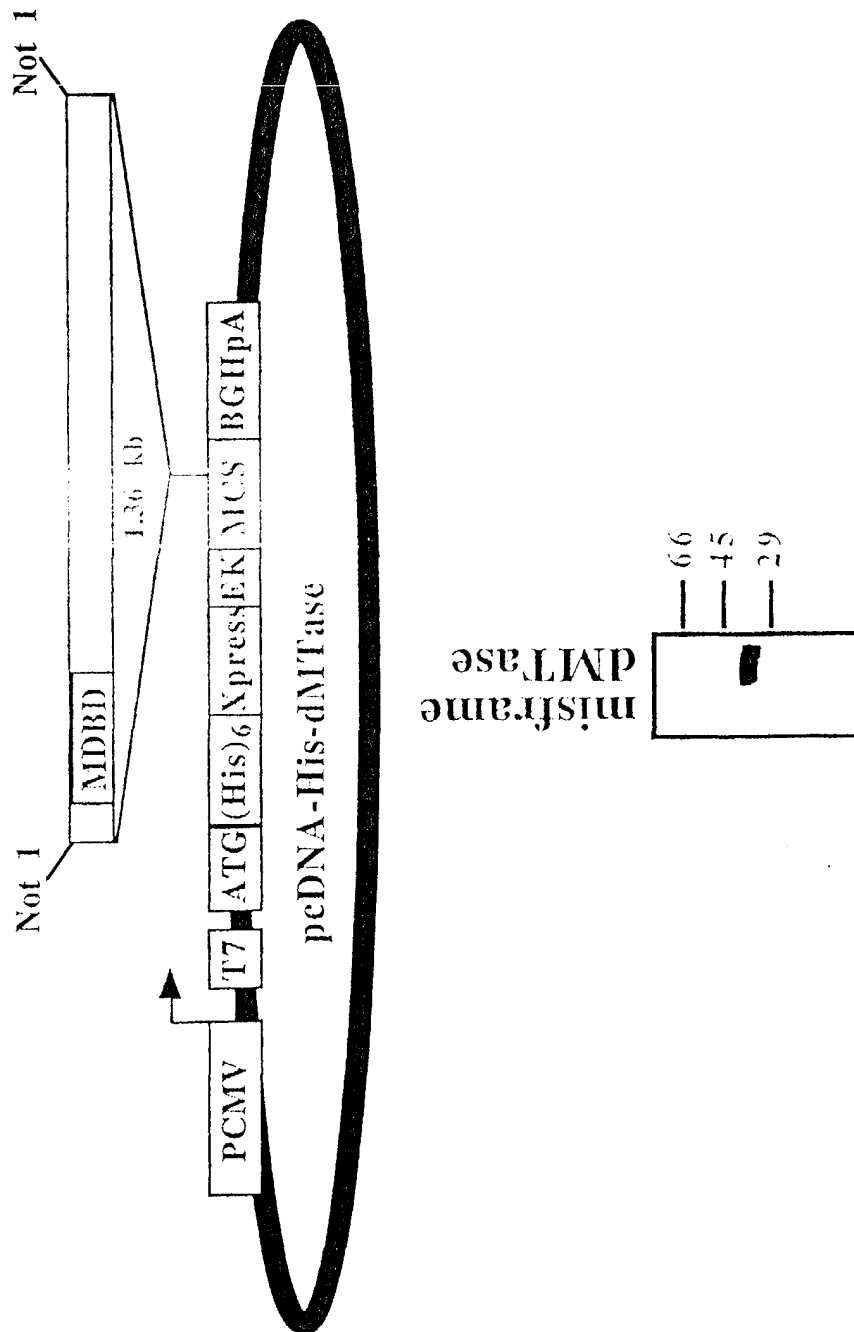
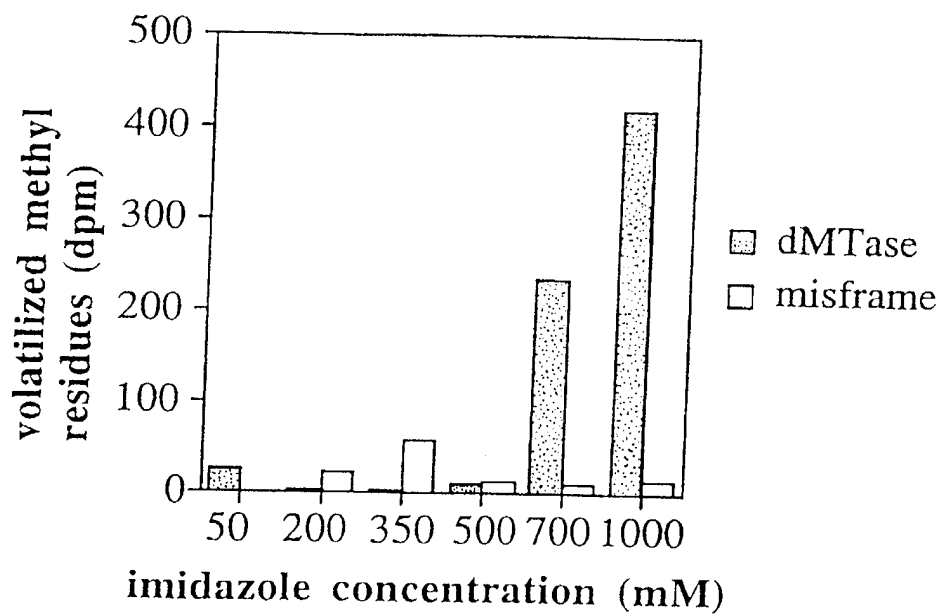
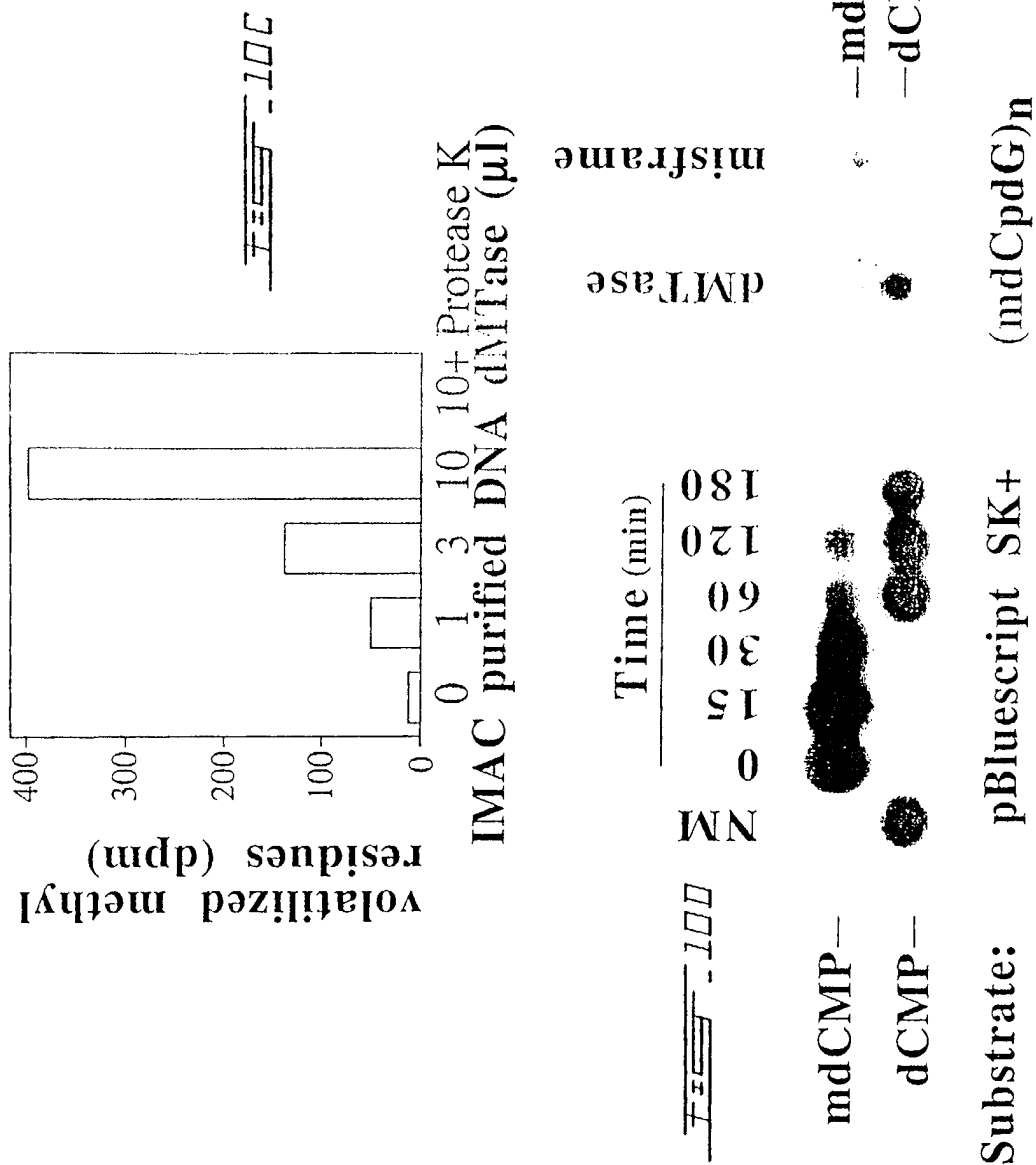


Fig. 10A

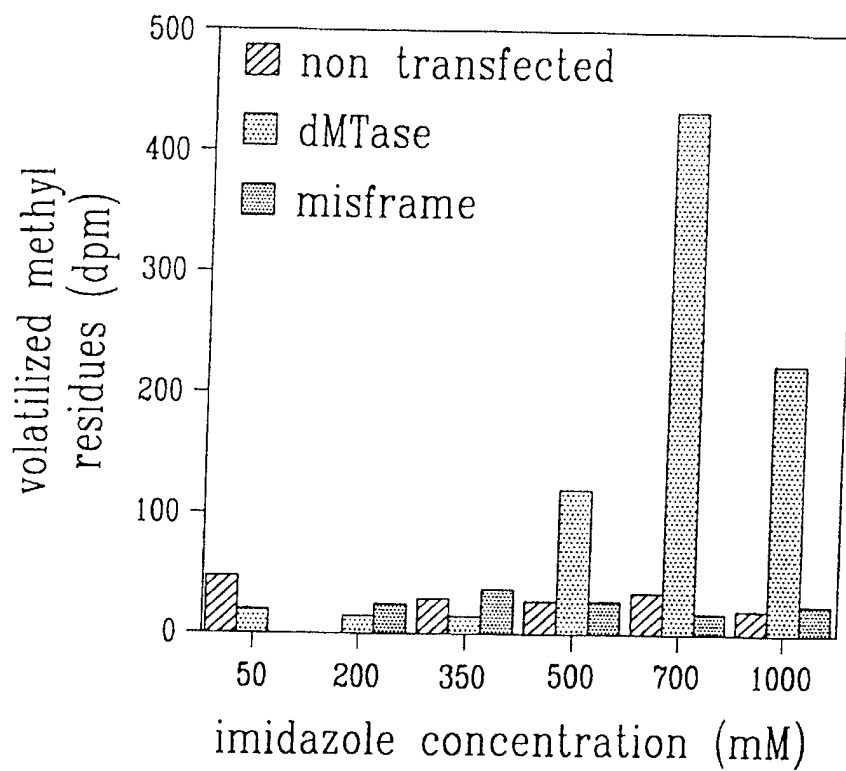
36/50

FIG. 10B

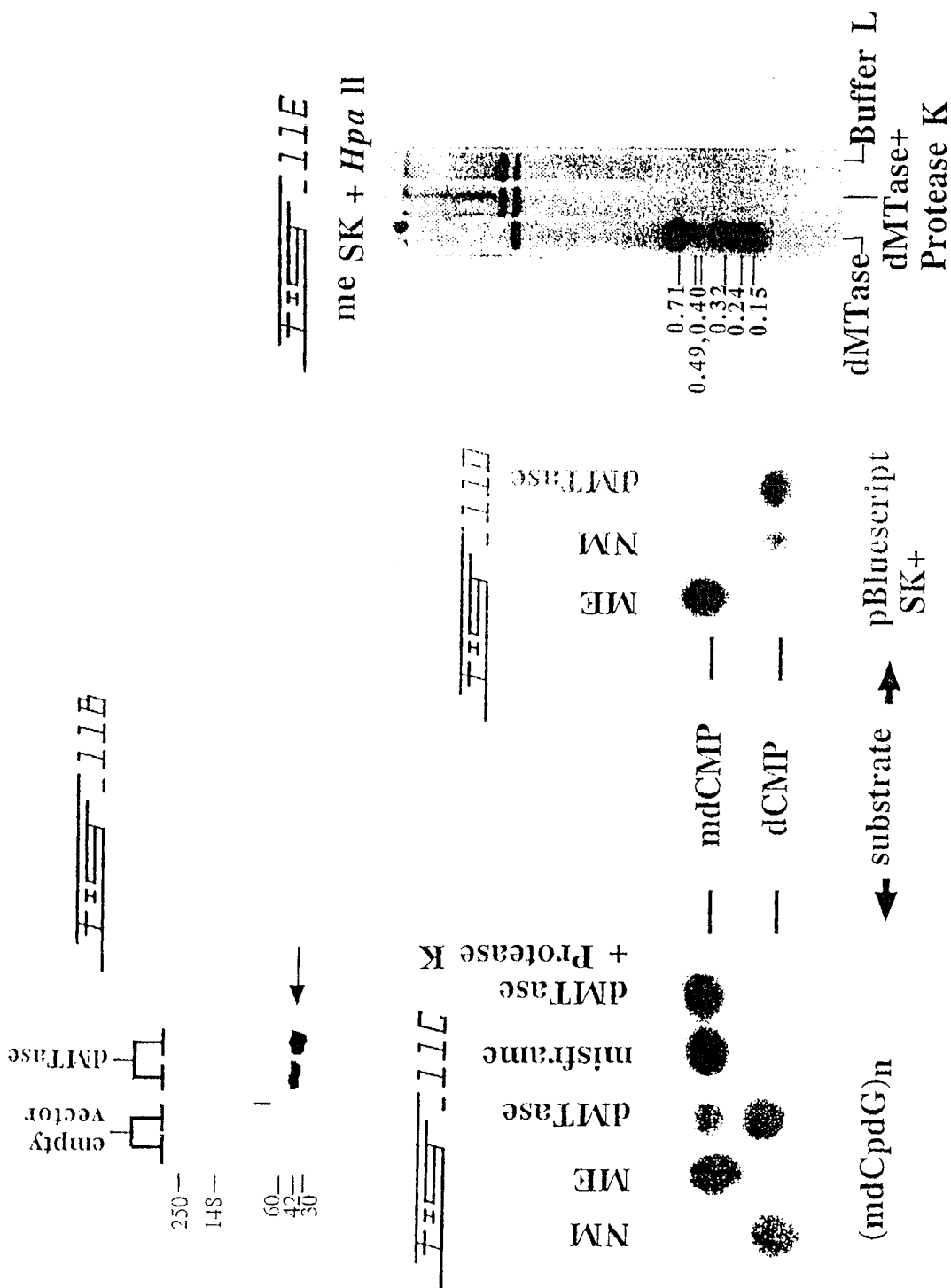
37/50



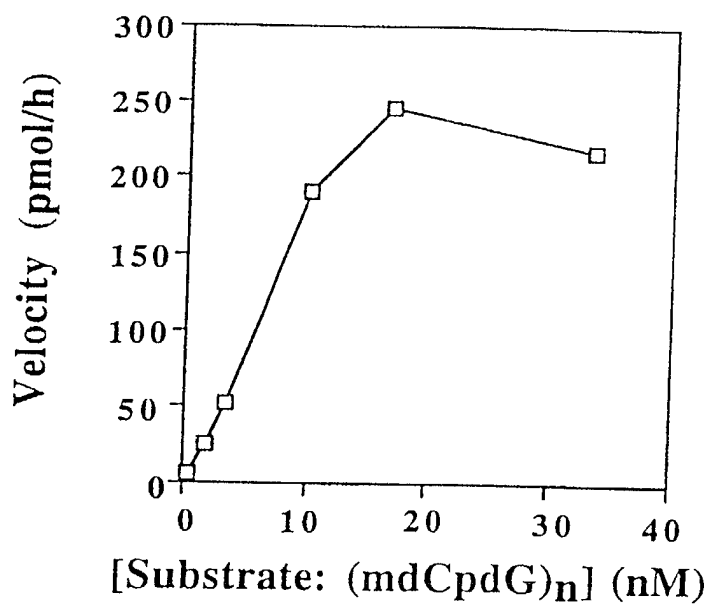
38/50

FIG. 11A

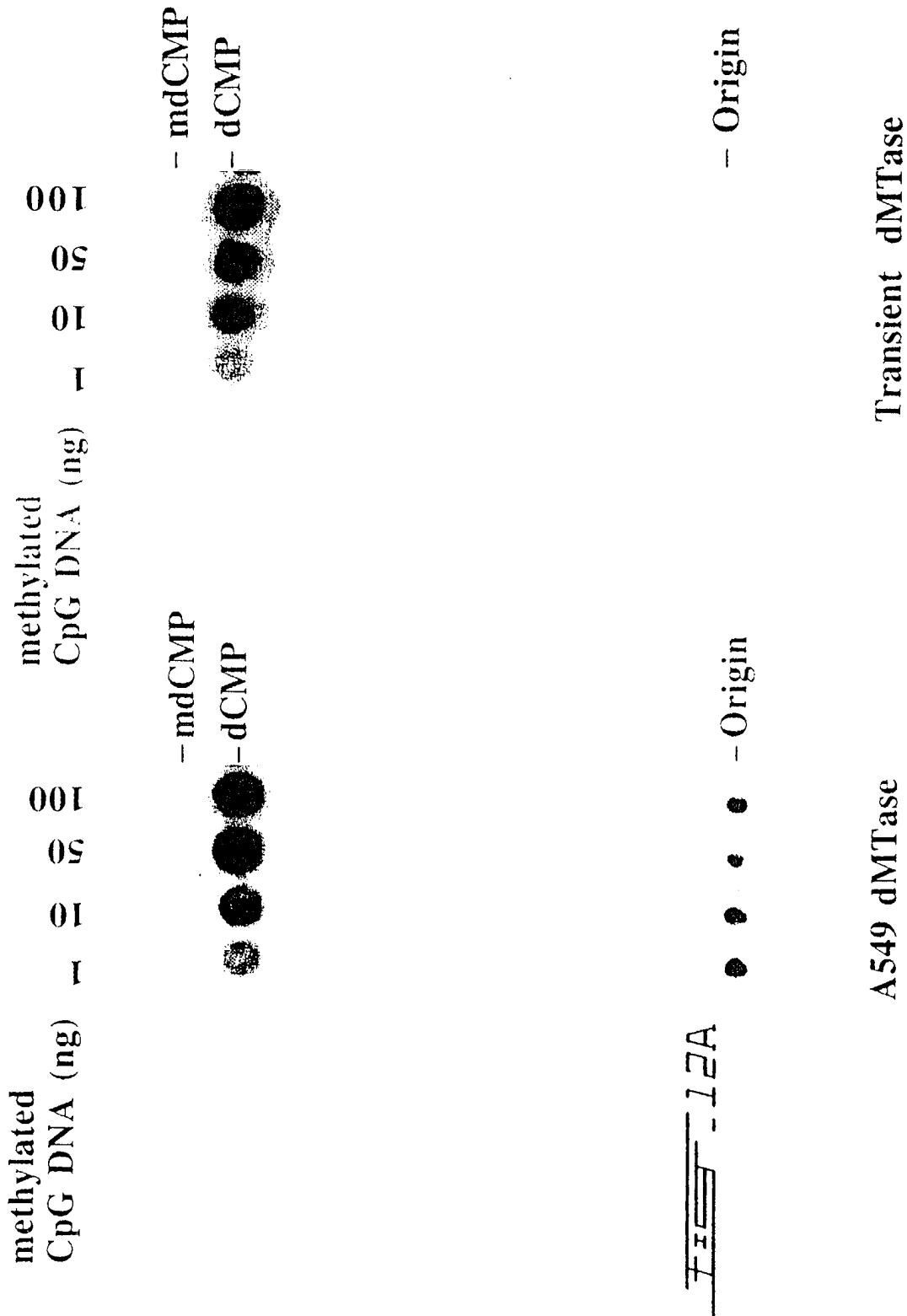
39/50



40/50

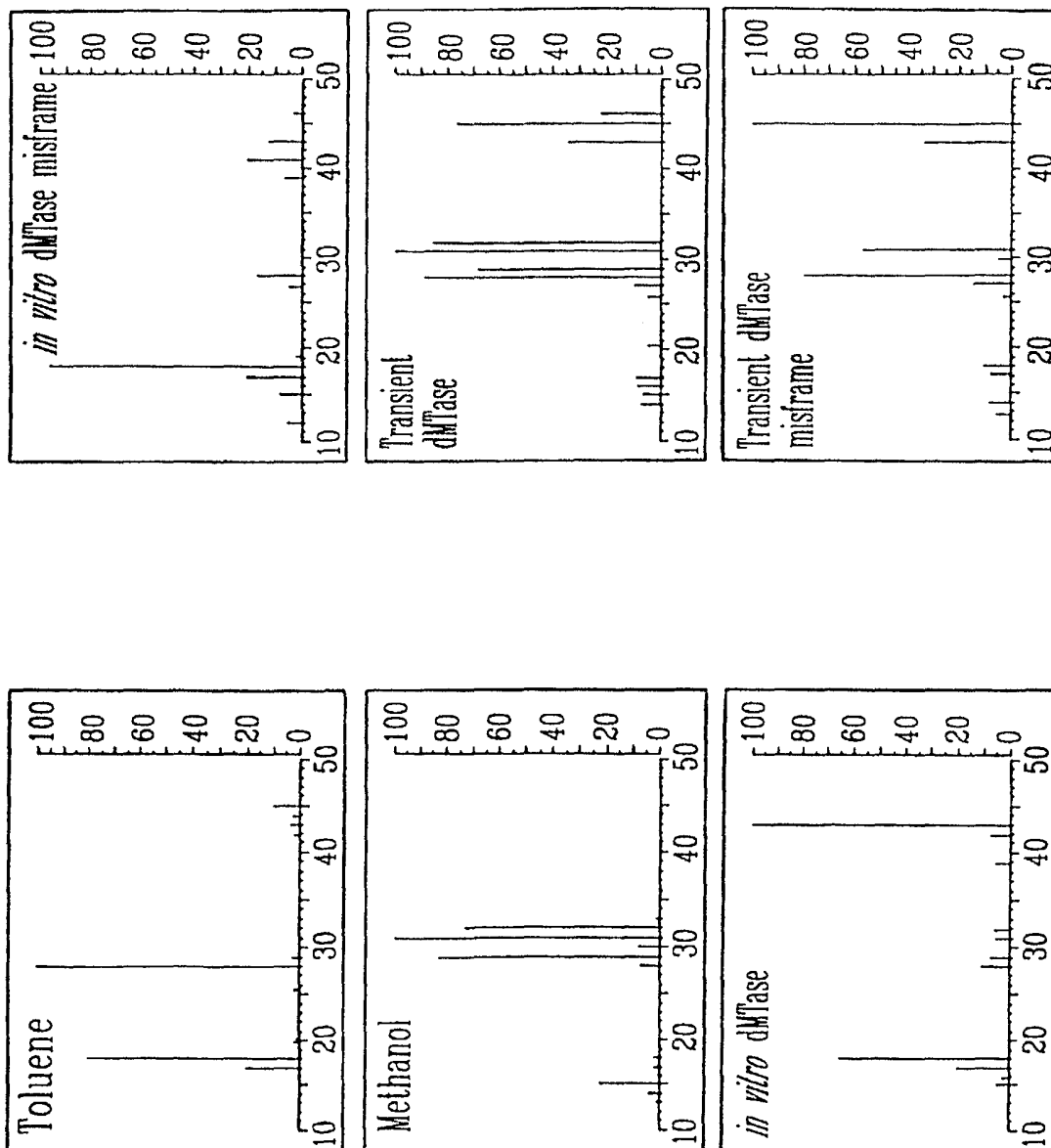
FIG. 11F

41/50



42/50

FIG - 12B



43/50

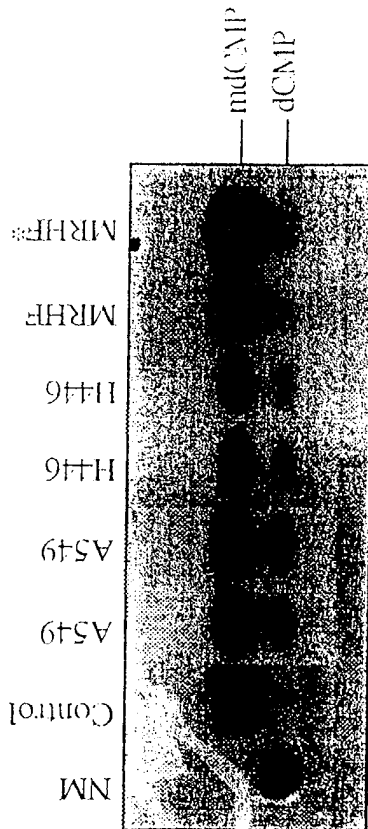
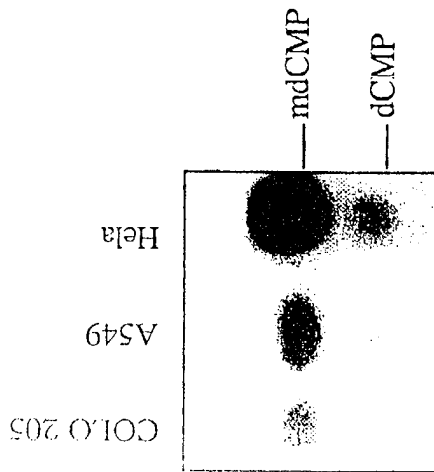


FIG. 13A

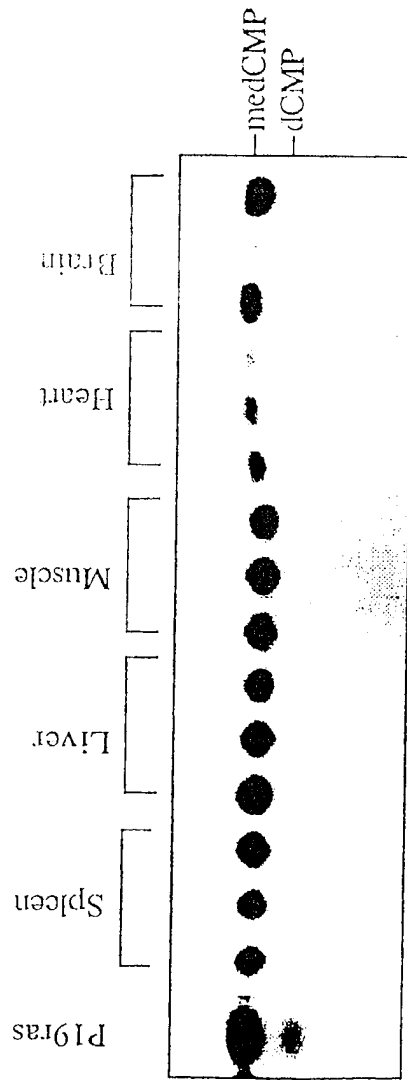
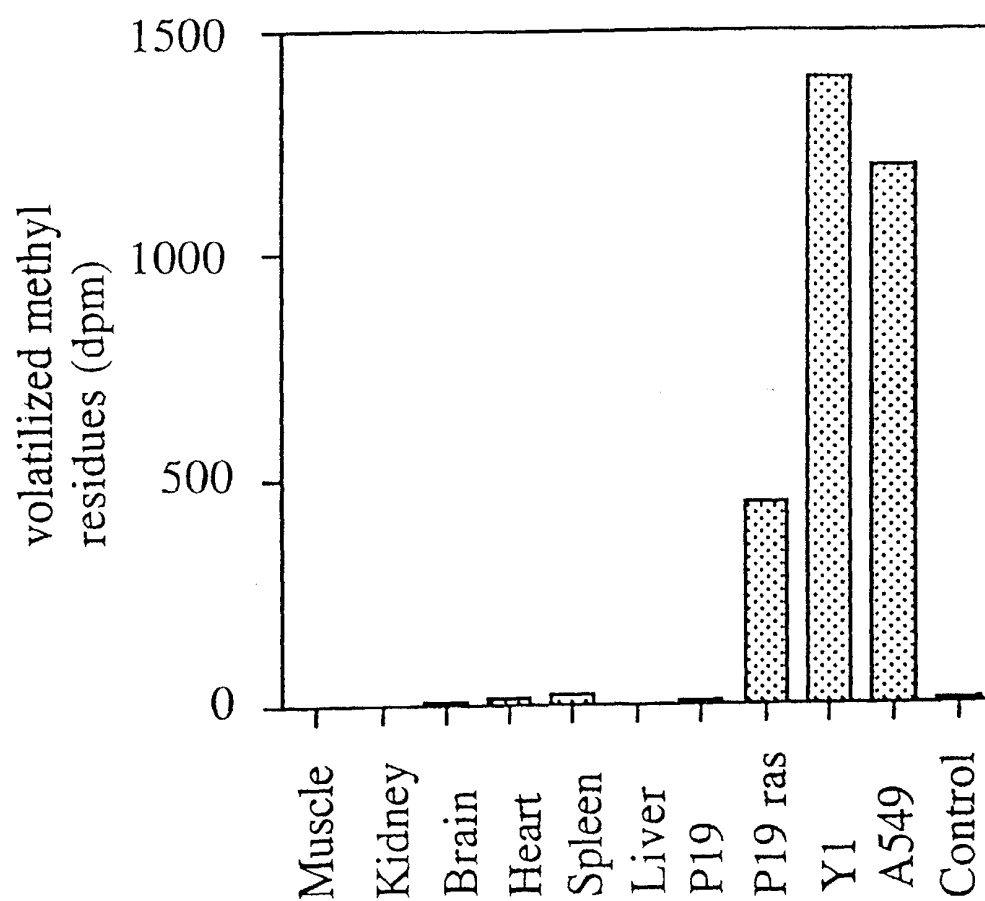


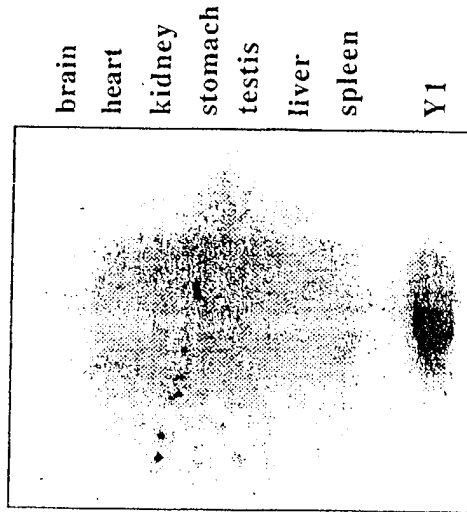
FIG. 13B

44/50

~~FIGURE 1~~ - 13C

45/50

dMTase



18s

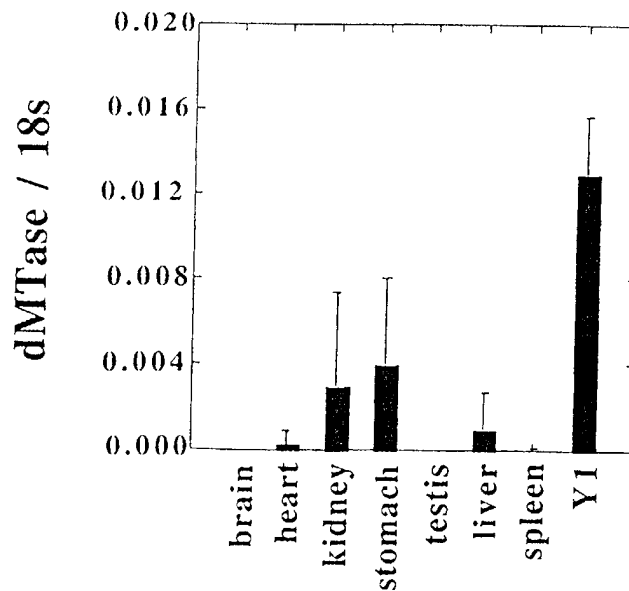
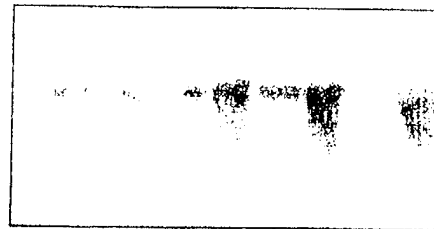
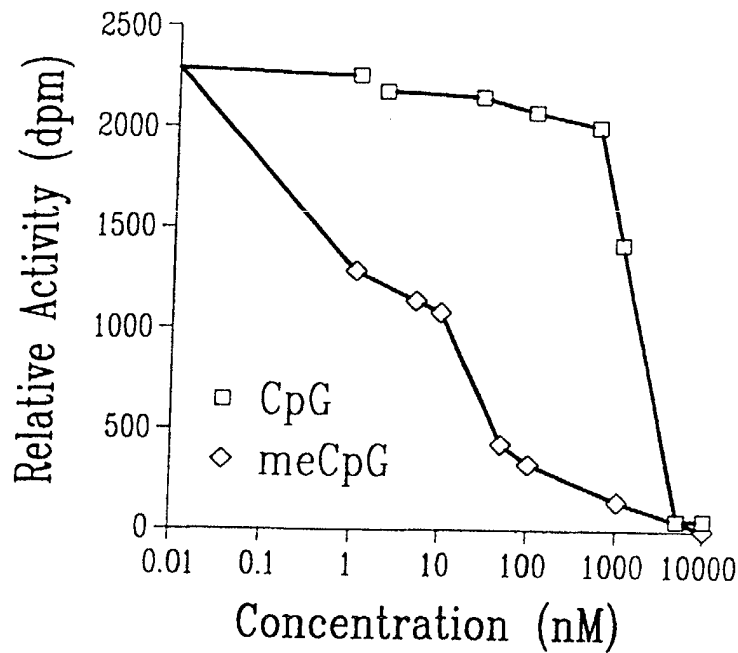
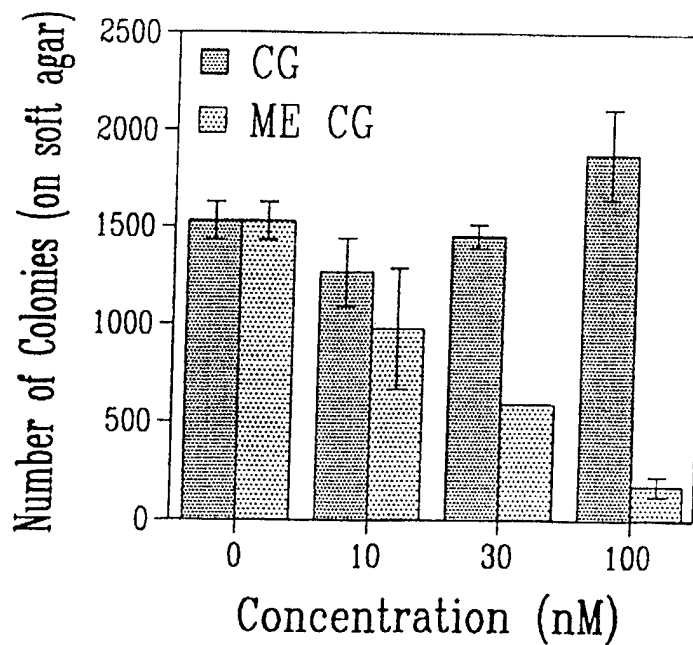


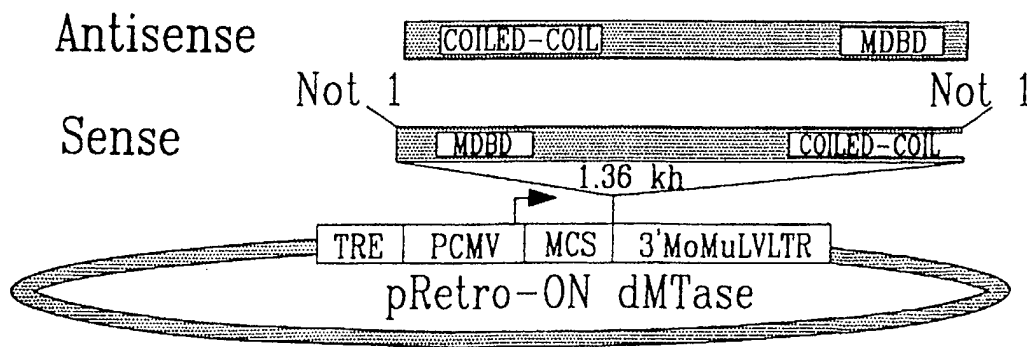
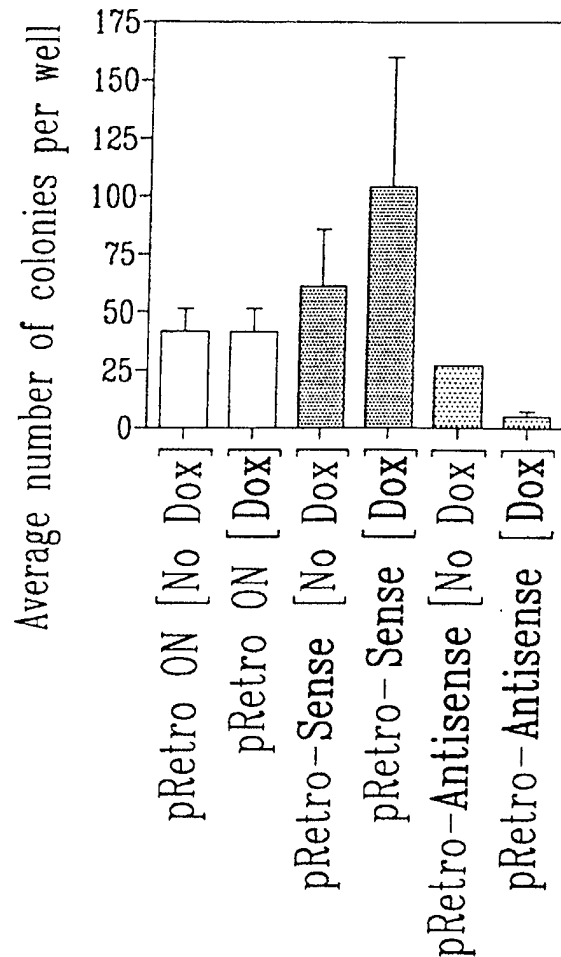
FIG. 130



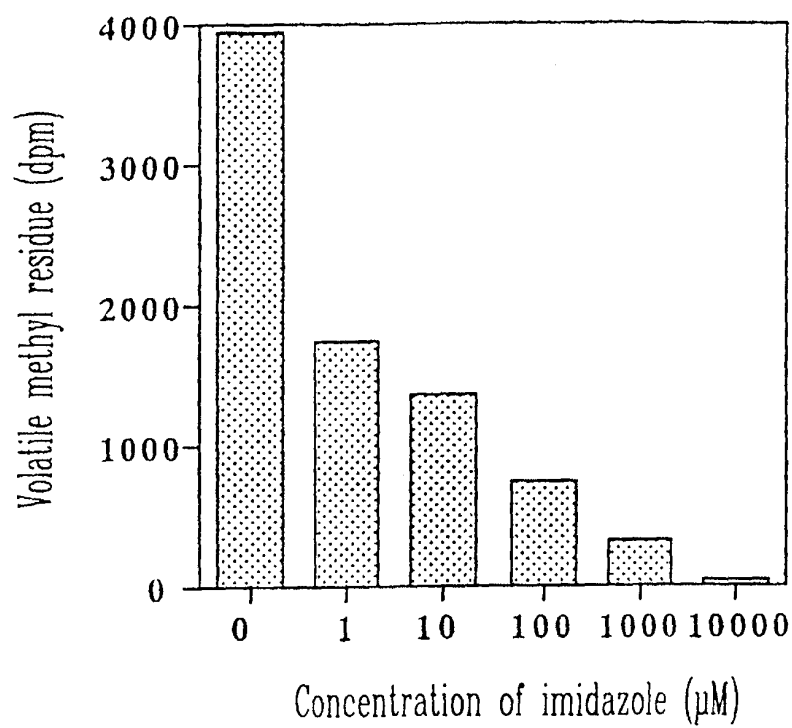
47/50

FIG. 14BFIG. 14C

48/50

FIG. 15

49/50

FIG. 16

50/50

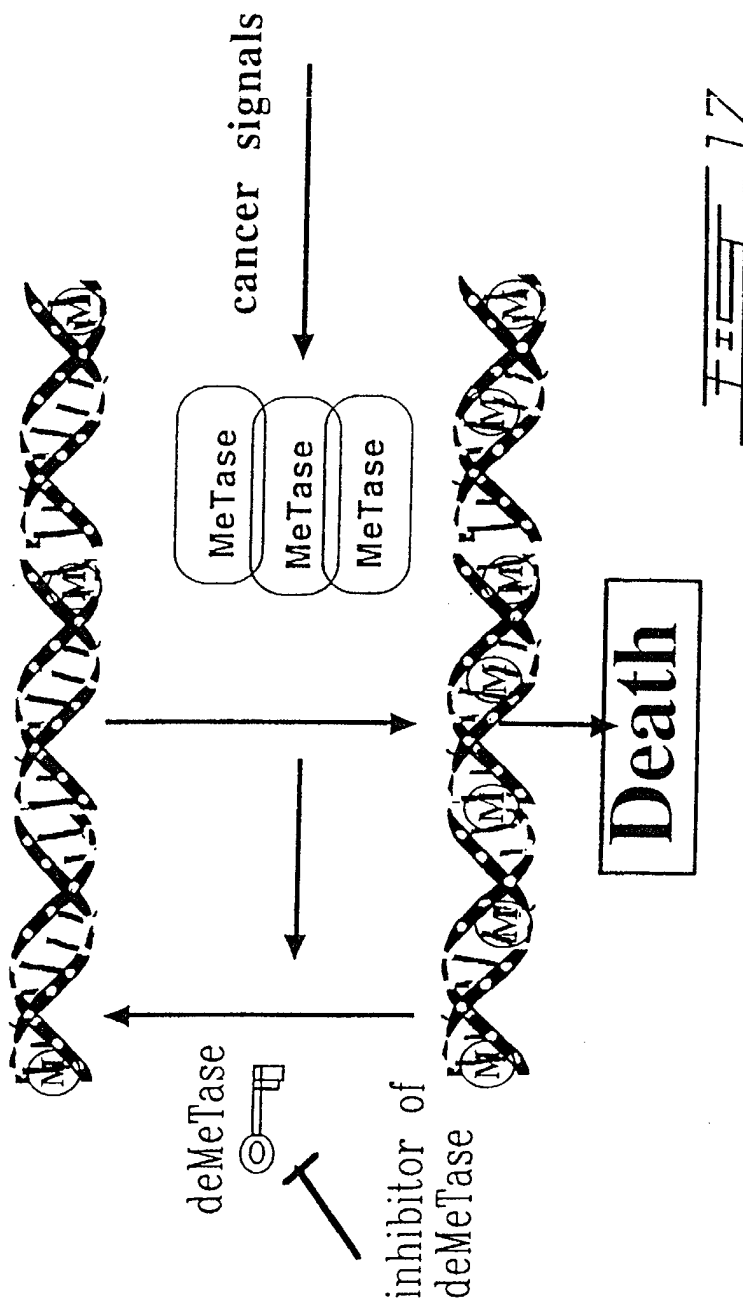


FIG. 17

SEQUENCE LISTING

<110> MCGILL UNIVERSITY
SZYF, Moshe
BHATTACHARYA, Sanjoy K.
RAMCHANDANI, Shyam

<120> DNA DEMETHYLASE, THERAPEUTIC AND
DIAGNOSTIC USES THEREOF

<130> 1770-183"PCT" FC/ld

<150> CA 2,220,805

<151> 1997-11-12

<150> CA 2,230,991

<151> 1998-05-11

<160> 10

<170> FastSEQ for Windows Version 3.0

<210> 1

<211> 1804

<212> DNA

<213> Unknown

<400> 1

| | | | | | | |
|------------|------------|------------|-------------|-------------|-------------|------|
| ccgctctgcg | ggcggggcgg | gtctccggga | ttccaagggc | toggttacgg | aagaagcgca | 60 |
| gagccggctg | gggagggggc | tggatgcgcg | cgcacccggg | gggaggccgc | tgctgcccgg | 120 |
| agcaggagga | gggggagagc | gcggcgggcg | gcagcggcgc | tggcggcgac | tccgccatag | 180 |
| agcagggggg | ccagggcagc | gcgctcgctc | cgtcccccgt | gagcggcgctg | cgcagggaag | 240 |
| gcgctcgggg | cggcggccgt | ggccgggggc | ggtggaagca | ggcggcccg | ggcggcggcg | 300 |
| tctgtggcgg | tggccgtggc | cgtggccggg | gtcggggccg | tggccggggc | cggggccggg | 360 |
| gccgcggcgg | tcccagagt | ggcggcagcg | gccttgccgg | cgcggcggcg | ggcggcggcg | 420 |
| ggggctgcgg | cgtcggcagc | ggtggcggcg | tcgccccccg | gcgggatcct | gtccctttcc | 480 |
| cgtcggggag | ctcggggccg | gggcccagg | gaccccgggc | cacggagagc | gggaagagga | 540 |
| tggactgcc | ggccctcccc | cccggatgga | agaaggagga | agtgatccga | aaatcagggc | 600 |
| tcagtgtctg | caagagcgat | gtctactact | tcagtccaag | tggtagaag | ttcagaagta | 660 |
| aacctcagct | ggcaagatac | ctgggaaatg | ctgttgacct | tagcagtttt | gacttcagga | 720 |
| ccggcaagat | gatgcctagt | aaattacaga | agaacaagca | gagactccgg | aatgaccccc | 780 |
| tcaatcagaa | caagggtaaa | ccagacctga | acacaacatt | gccaataga | caaactgcat | 840 |
| caattttcaa | gcaaccagta | accaaattca | cgaaccaccc | gagcaataag | gtgaagtcag | 900 |
| acccccagcg | gatgaatgaa | caaccacgtc | agctttttctg | ggagaagagg | ctacaaggac | 960 |
| ttagcgcac | agatgtaaca | gaacaaatta | taaaaaccat | ggagctacct | aaaggtcttc | 1020 |
| aaggagtccg | tccaggtagc | aatgacgaga | ccctctctgc | tgctgtggcc | agtgcctttac | 1080 |
| acacaagctc | tgcgcccac | acaggacaag | tctctgctgc | cgtggaaaag | aacctcgctg | 1140 |
| tttggtctaa | cacatctcaa | ccccctctga | aagctttcat | tgttacagat | gaagacatta | 1200 |
| ggaaacagga | agagcgagtc | caacaagtac | gcaagaaact | ggaggaggca | ctgatggccg | 1260 |
| acatcctgtc | ccgggctgcg | gacacggagg | aagtagacat | tgacatggac | agtggagatg | 1320 |
| aggcgtaaga | atatgatcag | gtaactttcg | actgaccttc | cccaagagca | aattgctaga | 1380 |
| aacagaatta | aaacatttcc | actgggtttc | gctgtgaaga | aaaagtgtac | ctgagcacat | 1440 |
| agctttttta | tagcactaac | caatgccttt | ttagatgtat | ttttgatgta | tatatctatt | 1500 |
| attccaaatg | atgtttat | tgaatcctag | gacttaaaat | gagtctttta | taatagcaag | 1560 |
| cagggccctt | ccggtgcagt | gcagctttga | ggccagggtg | agtctactgg | aaaggtagca | 1620 |
| cttacgtgaa | atatttgtt | ccccacagt | tttaataata | acagatcagg | agtaccaa | 1680 |

aagtttccca attaaagatt attatacttc actgtatata aacagatttt tatactttat 1740
 tgaaagaaga tacctgtaca ttcttccatc atcactgtaa agacaaataa atgactatat 1800
 tcac 1804

<210> 2
 <211> 411
 <212> PRT
 <213> Unknown

<400> 2
 Met Arg Ala His Pro Gly Gly Gly Arg Cys Cys Pro Glu Gln Glu Glu
 1 5 10 15
 Gly Glu Ser Ala Ala Gly Gly Ser Gly Ala Gly Gly Asp Ser Ala Ile
 20 25 30
 Glu Gln Gly Gly Gln Gly Ser Ala Leu Ala Pro Ser Pro Val Ser Gly
 35 40 45
 Val Arg Arg Glu Gly Ala Arg Gly Gly Gly Arg Gly Arg Gly Arg Trp
 50 55 60
 Lys Gln Ala Gly Arg Gly Gly Val Cys Gly Arg Gly Arg Gly Arg
 65 70 75 80
 Gly Arg Gly Arg Gly Arg Gly Arg Gly Arg Gly Arg Gly Arg Gly Arg
 85 90 95
 Pro Pro Ser Gly Gly Ser Gly Leu Gly Gly Asp Gly Gly Gly Cys Gly
 100 105 110
 Gly Gly Gly Ser Gly Gly Gly Gly Ala Pro Arg Arg Glu Pro Val Pro
 115 120 125
 Phe Pro Ser Gly Ser Ala Gly Pro Gly Pro Arg Gly Pro Arg Ala Thr
 130 135 140
 Glu Ser Gly Lys Arg Met Asp Cys Pro Ala Leu Pro Pro Gly Trp Lys
 145 150 155 160
 Lys Glu Glu Val Ile Arg Lys Ser Gly Leu Ser Ala Gly Lys Ser Asp
 165 170 175
 Val Tyr Tyr Phe Ser Pro Ser Gly Lys Lys Phe Arg Ser Lys Pro Gln
 180 185 190
 Leu Ala Arg Tyr Leu Gly Asn Thr Val Asp Leu Ser Ser Phe Asp Phe
 195 200 205
 Arg Thr Gly Lys Met Met Pro Ser Lys Leu Gln Lys Asn Lys Gln Arg
 210 215 220
 Leu Arg Asn Asp Pro Leu Asn Gln Asn Lys Gly Lys Pro Asp Leu Asn
 225 230 235 240
 Thr Thr Leu Pro Ile Arg Gln Thr Ala Ser Ile Phe Lys Gln Pro Val
 245 250 255
 Thr Lys Val Thr Asn His Pro Ser Asn Lys Val Lys Ser Asp Pro Gln
 260 265 270
 Arg Met Asn Glu Gln Pro Arg Gln Leu Phe Trp Glu Lys Arg Leu Gln
 275 280 285
 Gly Leu Ser Ala Ser Asp Val Thr Glu Gln Ile Ile Lys Thr Met Glu
 290 295 300
 Leu Pro Lys Gly Leu Gln Gly Val Gly Pro Gly Ser Asn Asp Glu Thr
 305 310 315 320
 Leu Leu Ser Ala Val Ala Ser Ala Leu His Thr Ser Ser Ala Pro Ile
 325 330 335
 Thr Gly Gln Val Ser Ala Ala Val Glu Lys Asn Pro Ala Val Trp Leu
 340 345 350
 Asn Thr Ser Gln Pro Leu Cys Lys Ala Phe Ile Val Thr Asp Glu Asp
 355 360 365

Ile Arg Lys Gln Glu Glu Arg Val Gln Gln Val Arg Lys Lys Leu Glu
370 375 380
Glu Ala Leu Met Ala Asp Ile Leu Ser Arg Ala Ala Asp Thr Glu Glu
385 390 395 400
Met Asp Ile Glu Met Asp Ser Gly Asp Glu Ala
405 410

<210> 3
<211> 1589
<212> DNA
<213> Unknown

<400> 3
cacgcgcggg cgggtgggcg gagcggcccc cctagcgggg gctgtgaagc gcggggaggg 60
ggccgagcgg gtggcgaagc cggcgcgcgc cgggctgggg gcggaggggc gagggccgtg 120
ggacagaaca gctgcgcgca gtggcggcgg cggagggagc cgaatcggcg acgagcccgg 180
gggtcgcaac ttgcagaagc ggccggcggc gcggcatcgg ccacggcggg cggaaaagcc 240
ggggcgcaat ggagcggaag aggtgggagt gcccggcgct cccgcagggc tgggaaaggg 300
aagaagtgcc caggaggtcg gggctgtcgg ccggccacag ggatgtcttt tactatagcc 360
ccagcgggaa gaagtccgc agcaagccac aactggcacg ttacctgggc ggatccatgg 420
acctcagcac cttcgacttc cgcaccgga agatgttgat gaacaagatg aataagagtc 480
gccagcgtgt gcgctatgat tcttccaacc aggtcaaggc caagcctgac ctgaacaccg 540
cgctgcctgt acggcagact gcacccatct tcaagcaacc ggtgaccaag atcaccaacc 600
accccagcaa caaggtcaag agcgaccgc agaaggcagt ggaccagccg aggcagcttt 660
tgtgggagaa gaagctaagt ggattgagt cctttgacat tgcagaagaa ctggtcagga 720
ccatggactt gcccaagggc ctgcaggag tgggccctgg ctgtacagat gagacgtgc 780
tgtcagccat tgcagtgct ctacacacca gcacctgcc cattacaggc cagctctctg 840
cagcgtgga gaagaacct ggtgtgtggc tgaacactgc acagccactg tgcaaagcct 900
tcattggtgac agatgacgac atcaggaagc aggaggagct ggtacagcag gtacgggaagc 960
gcttgaggaa ggactgatg gccgacatgc tagctcatgt ggaggagctt gcccagagac 1020
ggggggcacc actggacaag ccctgtgcag aggaggaaag ggaggaggaa gaggaggagg 1080
aagagccgga gccagagcga gtgtagcaca ggtgccctgc ccaagtctgg gctgcagact 1140
gccttcagcc ttgcctggac caggtagggg ccagacctgt aggaggcagc cgtccacctc 1200
ctttccaaag cctcctgctt ccaggtctca gtgcaggag cccctgtgga ccttgaactc 1260
acttgctcct gcgctgctg gcaggaagcc ccacactgaa agcagatgag cagtgaacca 1320
actgagaggc cacctggaca cagtcacctc ctgcctcct tatcatagga caaggccttg 1380
cttggcaccg aggagctggg agccgtgttg ggtgtgggag gaagtttctg gaaacacacc 1440
tggtatgcc caccttatgt ccctaaggct attacaggcc agggtttgga ctgctcggc 1500
ccacagggt gccagcctc ccacactga ggtcagcag cccaccagga agtcactttc 1560
cttcaataaa ctgatggtag gaacttgtg 1589

<210> 4
<211> 291
<212> PRT
<213> Unknown

<400> 4
Met Glu Arg Lys Arg Trp Glu Cys Pro Ala Leu Pro Gln Gly Trp Glu
1 5 10 15
Arg Glu Glu Val Pro Arg Arg Ser Gly Leu Ser Ala Gly His Arg Asp
20 25 30
Val Phe Tyr Tyr Ser Pro Ser Gly Lys Lys Phe Arg Ser Lys Pro Gln
35 40 45
Leu Ala Arg Tyr Leu Gly Gly Ser Met Asp Leu Ser Thr Phe Asp Phe
50 55 60
Arg Thr Gly Lys Met Leu Met Ser Lys Met Asn Lys Ser Arg Gln Arg
65 70 75 80

Val Arg Tyr Asp Ser Ser Asn Gln Val Lys Gly Lys Pro Asp Leu Asn
 85 90 95
 Thr Ala Leu Pro Val Arg Gln Thr Ala Ser Ile Phe Lys Gln Pro Val
 100 105 110
 Thr Lys Ile Thr Asn His Pro Ser Asn Lys Val Lys Ser Asp Pro Gln
 115 120 125
 Lys Ala Val Asp Gln Pro Arg Gln Leu Phe Trp Glu Lys Lys Leu Ser
 130 135 140
 Gly Leu Asn Ala Phe Asp Ile Ala Glu Glu Leu Val Lys Thr Met Asp
 145 150 155 160
 Leu Pro Lys Gly Leu Gln Gly Val Gly Pro Gly Cys Thr Asp Glu Thr
 165 170 175
 Leu Leu Ser Ala Ile Ala Ser Ala Leu His Thr Ser Thr Met Pro Ile
 180 185 190
 Thr Gly Gln Leu Ser Ala Ala Val Glu Lys Asn Pro Gly Val Trp Leu
 195 200 205
 Asn Thr Thr Gln Pro Leu Cys Lys Ala Phe Met Val Thr Asp Glu Asp
 210 215 220
 Ile Arg Lys Gln Glu Glu Leu Val Gln Gln Val Arg Lys Arg Leu Glu
 225 230 235 240
 Glu Ala Leu Met Ala Asp Met Leu Ala His Val Glu Glu Leu Ala Arg
 245 250 255
 Asp Gly Glu Ala Pro Leu Asp Lys Ala Cys Ala Glu Asp Asp Glu
 260 265 270
 Glu Asp Glu Glu Glu Glu Glu Glu Glu Pro Asp Pro Asp Pro Glu Met
 275 280 285
 Glu His Val
 290

<210> 5
 <211> 1966
 <212> DNA
 <213> Unknown

<400> 5

| | | | | | | |
|------------|------------|-------------|-------------|------------|-------------|------|
| gggggcggtg | ccccgagaag | gcggagacaa | gatggccgcc | catagcgctt | ggaggaccta | 60 |
| agaggcggtg | gccggggcca | cgccccgggc | aggaggggcg | ctctgtgcgc | gcccgcctcta | 120 |
| tgatgcttgc | gcgcgtcccc | cgcgcgccgc | gctgcggggcg | gggcggtct | ccgggattcc | 180 |
| aagggtctcg | ttacggaaga | agcgagcgcg | cggctgggga | gggggctgga | tgcgcgcgca | 240 |
| ccggggggga | ggcgctgct | gcccggagca | ggaggagggg | gagagtgcgg | cgggcggcag | 300 |
| cggcgtggc | ggcgactccg | ccatagagca | ggggggccag | ggcagcgcg | tcgccccgtc | 360 |
| cccggtgagc | ggcggtgcga | gggaaggcgc | tcggggcggc | ggcgtggcc | gggggctgtg | 420 |
| gaagcaggcg | ggcgggggcg | gcggcgctctg | tgccgtggc | cggggcccgg | gcccgtggccg | 480 |
| gggacgggga | cgggggccgg | gccggggccg | cgccgtccc | ccgagtggcg | gcagcggcct | 540 |
| tggcgcgac | ggcgggcggt | gcggcgccgg | cggcagcggt | ggcgcgggcg | ccccccggcg | 600 |
| ggagccggtc | cctttcccg | cggggagcgc | ggggccgggg | cccaggggac | cccggggccac | 660 |
| ggagagcggg | aagaggatgg | attgcccggc | cctccccccc | ggatggaaga | aggagggaagt | 720 |
| gatccgaaaa | tctgggctaa | gtgctggcaa | gagcgatgtc | tactacttca | gtccaagtgg | 780 |
| taagaagtgc | agaagcaagc | ctcagttggc | aaggtacctg | ggaaatactg | ttgatctcag | 840 |
| cagttttgac | ttcagaactg | gaaagatgat | gcctagtaaa | ttacagaaga | acaaacagag | 900 |
| actgcgaac | gatcctctca | atcaaaataa | gggtaaacca | gacttgaata | caacattgcc | 960 |
| aattagacaa | acagcatcaa | ttttcaaaca | accggttaacc | aaagtcaaca | atcatcctag | 1020 |
| taataaagtg | aaatcagacc | cacaacgaat | gaatgaacag | ccacgtcagc | ttttctggga | 1080 |
| gaagaggcta | caaggactta | gtgcatacaga | tgtaacagaa | caaattataa | aaaccatgga | 1140 |
| actaccacaa | ggtcttcaag | gagttggtcc | aggtagcaat | gatgagaccc | ttttatctgc | 1200 |
| tggtgccagt | gctttgcaca | caagctctgc | gccaatcaca | gggcaagtct | ccgctgctgt | 1260 |
| ggaaaagaac | cctgctgttt | ggcttaacac | atctcaaccc | ctctgcaaag | cttttattgt | 1320 |

```

cacagatgaa gacatcagga aacaggaaga gcgagtacag caagtacgca agaaattgga 1380
agaagcactg atggcgagaca tcttgtcgcg agctgctgat acagaagaga tggatattga 1440
aatggacagt ggagatgaag cctaagaata tgatcaggta actttcgacc gactttcccc 1500
aagrgaaaat tctagaaaat tgaacaaaaa tgttccact ggcttttgcc tgtaagaaaa 1560
aaaatgtacc cgagcacata gagcttttta atagcactaa ccaatgcctt tttagatgta 1620
tttttgatgt atatatctat tattcaaaaa atcatgttta ttttgagtec taggacttaa 1680
aattagtctt ttgtaatatc aagcaggacc ctaagatgaa gctgagcttt tgatgccagg 1740
tgcaatctac tggaaatgta gcacttacgt aaaacatttg tttccccac agttttaata 1800
agaacagatc aggaattcta aataaatttc ccagttaaag attattgtga cttcactgta 1860
tataaacata tttttatact ttattgaaag gggacacctg tacattcttc catcatcact 1920
gtaaagacaa ataatgatt atattcacia aaaaaaaaaa aaaaaa 1960

```

<210> 6
 <211> 414
 <212> PRT
 <213> Unknown

<400> 6

```

Met Arg Ala His Pro Gly Gly Gly Arg Cys Cys Pro Glu Gln Glu Glu
 1           5           10           15
Gly Glu Ser Ala Ala Gly Gly Ser Gly Ala Gly Gly Asp Ser Ala Ile
          20          25          30
Glu Gln Gly Gly Gln Gly Ser Ala Leu Ala Pro Ser Pro Val Ser Gly
          35          40          45
Val Arg Arg Glu Gly Ala Arg Gly Gly Gly Arg Gly Arg Gly Arg Trp
          50          55          60
Lys Gln Ala Ala Arg Gly Gly Gly Val Cys Gly Arg Gly Arg Gly Arg
65          70          75          80
Gly Arg Gly Arg Gly Arg Gly Arg Gly Arg Gly Arg Gly Arg Gly Arg
          85          90          95
Pro Gln Ser Gly Gly Ser Gly Leu Gly Gly Asp Gly Gly Gly Gly Ala
          100         105         110
Gly Gly Cys Gly Val Gly Ser Gly Gly Gly Val Ala Pro Arg Arg Asp
          115         120         125
Pro Val Pro Phe Pro Ser Gly Ser Ser Gly Pro Gly Pro Arg Gly Pro
          130         135         140
Arg Ala Thr Glu Ser Gly Lys Arg Met Asp Cys Pro Ala Leu Pro Pro
          145         150         155         160
Gly Trp Lys Lys Glu Glu Val Ile Arg Lys Ser Gly Leu Ser Ala Gly
          165         170         175
Lys Ser Asp Val Tyr Tyr Phe Ser Pro Ser Gly Lys Lys Phe Arg Ser
          180         185         190
Lys Pro Gln Leu Ala Arg Tyr Leu Gly Asn Ala Val Asp Leu Ser Ser
          195         200         205
Phe Asp Phe Arg Thr Gly Lys Met Met Pro Ser Lys Leu Gln Lys Asn
          210         215         220
Lys Gln Arg Leu Arg Asn Asp Pro Leu Asn Gln Asn Lys Gly Lys Pro
          225         230         235         240
Asp Leu Asn Thr Thr Leu Pro Ile Arg Gln Thr Ala Ser Ile Phe Lys
          245         250         255
Gln Pro Val Thr Lys Phe Thr Asn His Pro Ser Asn Lys Val Lys Ser
          260         265         270
Asp Pro Gln Arg Met Asn Glu Gln Pro Arg Gln Leu Phe Trp Glu Lys
          275         280         285
Arg Leu Gln Gly Leu Ser Ala Ser Asp Val Thr Glu Gln Ile Ile Lys
          290         295         300

```

Thr Met Glu Leu Pro Lys Gly Leu Gln Gly Val Gly Pro Gly Ser Asn
 305 310 315 320
 Asp Glu Thr Leu Leu Ser Ala Val Ala Ser Ala Leu His Thr Ser Ser
 325 330 335
 Ala Pro Ile Thr Gly Gln Val Ser Ala Ala Val Glu Lys Asn Pro Ala
 340 345 350
 Val Trp Leu Asn Thr Ser Gln Pro Leu Cys Lys Ala Phe Ile Val Thr
 355 360 365
 Asp Glu Asp Ile Arg Lys Gln Glu Glu Arg Val Gln Gln Val Arg Lys
 370 375 380
 Lys Leu Glu Glu Ala Leu Met Ala Asp Ile Leu Ser Arg Ala Ala Asp
 385 390 395 400
 Thr Glu Glu Val Asp Ile Asp Met Asp Ser Gly Asp Glu Ala
 405 410

<210> 7
 <211> 2392
 <212> DNA
 <213> Unknown

<400> 7

| | | | | | | |
|-------------|------------|------------|-------------|------------|------------|------|
| agcgggcccga | ggagccggg | gcaatggagc | ggaagaggtg | ggagtgcccg | gcgctcccgc | 60 |
| agggctggga | gaggaagaa | gtgccagaa | ggcggggct | gtcggccggc | cacagggatg | 120 |
| tcttttacta | tagcccgagc | gggaagaagt | tccgcagcaa | gccgcagctg | gcgcgctacc | 180 |
| tgggcggctc | catggacctg | agcaccctcg | acttccgcac | gggcaagatg | ctgatgagca | 240 |
| agatgaacaa | gagccgccag | cgcgtgcgct | acgactcctc | caaccaggtc | aagggaagc | 300 |
| ccgacctgaa | cacggcgctg | cccgctgcgc | agacggcgctc | catcttcaag | cagccggtga | 360 |
| ccaagattac | caaccacccc | agcaacaagg | tcaagagcga | cccgcagaag | gcggtggacc | 420 |
| agccgcgcga | gctcttcttg | gagaagaagc | tgagcggcct | gaacgccttc | gacattgctg | 480 |
| aggagctggg | caagaccatg | gacctcccca | agggcctgca | gggggtggga | cctggctgca | 540 |
| cggatgagac | gctgctgtcg | gccatgcgca | gcgcctgca | cactagcacc | atgcccata | 600 |
| cgggacagct | ctcggccgcc | gtggagaaga | acccggcgct | atggctcaac | accacgcagc | 660 |
| ccctgtgcaa | agccttcctg | gtgaccgacg | aggacatcag | gaagcaggaa | gagctgggtg | 720 |
| agcagggtcg | gaagcggctg | gaggaggcgc | tgatggccga | catgctggcg | cacgtggagg | 780 |
| agctggcccg | tgacggggag | gcgcgctg | acaaggcctg | cgctgaggac | gacgacgagg | 840 |
| aagacgagga | ggaggaggag | gaggagcccc | acccggaccc | ggagatggag | cacgtctagg | 900 |
| gcagaggccc | tgccgagagc | ccgtgctgcc | tgctggagcc | gcctgcagac | gcggtcctcg | 960 |
| gccccacgtg | aaccaggctc | ggcggcgaag | cccagccttg | gagacacca | ggaggaaggc | 1020 |
| cgtgctcctg | gctccctcct | cggcccgctc | ccactcccg | gggcctcggg | gcacacagct | 1080 |
| ggggctgccc | ccaccgaaa | gacctccac | gctcgtcctc | tacagagtcc | ggcttcggga | 1140 |
| agtgcgggtg | gctcctgggc | cctgctggc | tccttacgac | ctttgggctc | gaggccagct | 1200 |
| cctccccatg | cccgctgtcc | cagctccttg | agactggaga | gcagccagca | ggtgcccggc | 1260 |
| agctcggcgc | cacggcttgc | tgacagctgg | gagggtttct | cggtctggag | gcgtagtgtt | 1320 |
| gaaactcaca | tcacccactg | tgacagctga | ggacgggact | ctggtctgct | gtggggggca | 1380 |
| tgcaggacgg | cgccactctc | tgccctgcca | tgcggtgggt | ggtgccacag | agcctcaccg | 1440 |
| tgccctgagtg | gcgtgcccag | ggaggccgct | ctccttcagt | aaatgtaaca | cagtcgaggc | 1500 |
| acgtcatcgg | gcagccttcc | ctgtgtgcca | acgcacgctc | tcgcttctga | aaaccaaact | 1560 |
| ccagccgctg | ccagtcggga | cttggtcgcc | cggcgtgcc | agaatgctcc | actgccagcc | 1620 |
| ggccccctg | cctcggtttc | ccttctgttt | agtggcgaca | caggcaccca | gctttggggg | 1680 |
| ggtgctgacg | ctcccagggg | tgccaggagc | cactgggaca | gggtgaggct | cccagacgct | 1740 |
| cctcgagggtg | cccagctctc | cagggagctt | ctggcccaag | gcgttcttga | gggatctgct | 1800 |
| ccttaacccc | ccagtgcctt | ggcgagggca | ggttccaagc | cacagacgcc | tgccccgagt | 1860 |
| ggactttgcg | gccagtcctt | gggtgccttc | ctgggcccctg | cttgcccagt | gagggttctc | 1920 |
| aacgggtggg | ttcawtgggc | tggcccvagc | ggcccccac | ctgcattgac | cttagggcca | 1980 |
| tagagagggc | ctgtcccggg | gtgccccag | cgaaggatct | ggtcgctgcc | ccagggggac | 2040 |
| tgatgggcaa | gagtcgcccc | tgtggctgga | ctgtgaccat | ccctgatggg | gcctgaccgc | 2100 |
| gggagctgag | gaagcgccgc | tccaccgtct | gcctccaag | gaccgcagtg | gaggcagtg | 2160 |

```

gctggcagct tctgctgct cctgtcaga gtcaaagcac aaatcctcag gacgggctca 2220
agggccaggg cagccgaggg aagctccagg tggggaccac gtcttctga ggttggtgcc 2280
cactggctgg gaccctttgc agtggggtgg cctccctct gtctgcctgg tggaggagc 2340
cgtgggcgtg gggacgtgac tgaataaagc caccatgggt ggatgtgctt gg 2392

```

<210> 8
 <211> 285
 <212> PRT
 <213> Unknown

<400> 8

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Glu | Arg | Lys | Arg | Trp | Glu | Cys | Pro | Ala | Leu | Pro | Gln | Gly | Trp | Glu |
| 1 | | | 5 | | | | | | 10 | | | | | 15 | |
| Arg | Glu | Glu | Val | Pro | Arg | Arg | Ser | Gly | Leu | Ser | Ala | Gly | His | Arg | Asp |
| | | | 20 | | | | | 25 | | | | | 30 | | |
| Val | Phe | Tyr | Ser | Pro | Ser | Gly | Lys | Lys | Phe | Arg | Ser | Lys | Pro | Gln | |
| | | 35 | | | | 40 | | | | | 45 | | | | |
| Leu | Ala | Arg | Tyr | Leu | Gly | Gly | Ser | Met | Asp | Leu | Ser | Thr | Phe | Asp | Phe |
| | | 50 | | | 55 | | | | | 60 | | | | | |
| Arg | Thr | Gly | Lys | Met | Leu | Met | Asn | Lys | Met | Asn | Lys | Ser | Arg | Gln | Arg |
| 65 | | | | 70 | | | | | 75 | | | | | 80 | |
| Val | Arg | Tyr | Asp | Ser | Ser | Asn | Gln | Val | Lys | Gly | Lys | Pro | Asp | Leu | Asn |
| | | | 85 | | | | | 90 | | | | | | 95 | |
| Thr | Ala | Leu | Pro | Val | Arg | Gln | Thr | Ala | Ser | Ile | Phe | Lys | Gln | Pro | Val |
| | | | 100 | | | | | 105 | | | | | 110 | | |
| Thr | Lys | Ile | Thr | Asn | His | Pro | Ser | Asn | Lys | Val | Lys | Ser | Asp | Pro | Gln |
| | | 115 | | | | 120 | | | | | | 125 | | | |
| Lys | Ala | Val | Asp | Gln | Pro | Arg | Gln | Leu | Phe | Trp | Glu | Lys | Lys | Leu | Ser |
| | | 130 | | | | 135 | | | | | 140 | | | | |
| Gly | Leu | Ser | Ala | Phe | Asp | Ile | Ala | Glu | Glu | Leu | Val | Arg | Thr | Met | Asp |
| 145 | | | | 150 | | | | | 155 | | | | | 160 | |
| Leu | Pro | Lys | Gly | Leu | Gln | Gly | Val | Gly | Pro | Gly | Cys | Thr | Asp | Glu | Thr |
| | | | 165 | | | | | 170 | | | | | 175 | | |
| Leu | Leu | Ser | Ala | Ile | Ala | Ser | Ala | Leu | His | Thr | Ser | Thr | Leu | Pro | Ile |
| | | 180 | | | | | | 185 | | | | | 190 | | |
| Thr | Gly | Gln | Leu | Ser | Ala | Ala | Val | Glu | Lys | Asn | Pro | Gly | Val | Trp | Leu |
| | | 195 | | | | | 200 | | | | | 205 | | | |
| Asn | Thr | Ala | Gln | Pro | Leu | Cys | Lys | Ala | Phe | Met | Val | Thr | Asp | Asp | Asp |
| 210 | | | | | 215 | | | | | | 220 | | | | |
| Ile | Arg | Lys | Gln | Glu | Glu | Leu | Val | Gln | Gln | Val | Arg | Lys | Arg | Leu | Glu |
| 225 | | | | 230 | | | | | 235 | | | | | 240 | |
| Glu | Ala | Leu | Met | Ala | Asp | Met | Leu | Ala | His | Val | Glu | Glu | Leu | Ala | Arg |
| | | | 245 | | | | | 250 | | | | | 255 | | |
| Asp | Gly | Glu | Ala | Pro | Leu | Asp | Lys | Ala | Cys | Ala | Glu | Glu | Glu | Glu | Glu |
| | | 260 | | | | | 265 | | | | | | 270 | | |
| Glu | Glu | Glu | Glu | Glu | Glu | Glu | Pro | Glu | Pro | Glu | Arg | Val | | | |
| | | 275 | | | | | 280 | | | | | 285 | | | |

<210> 9
 <211> 17
 <212> DNA
 <213> Unknown

<400> 9
 ctggcaagag cgatgtc

WO 99/24583

8/8

PCT/CA98/01059

<210> 10
<211> 22
<212> DNA
<213> Unknown

<400> 10
agttctgggttt acccttatatt tg

22

INTERNATIONAL SEARCH REPORT

national Application No
PCT/CA 98/01059

| A. CLASSIFICATION OF SUBJECT MATTER IPC 6 C12N15/52 C12N9/00 A61K38/43 C07K16/40 C12N15/11 A61K48/00 G01N33/50 | | |
|---|--|--|
| According to International Patent Classification (IPC) or to both national classification and IPC | | |
| B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC 6 C12N A61K G01N | | |
| Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched | | |
| Electronic data base consulted during the international search (name of data base and, where practical, search terms used) | | |
| C. DOCUMENTS CONSIDERED TO BE RELEVANT | | |
| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
| T | BHATTACHARYA SK ET AL: "A mammalian protein with specific demethylase activity for mCpG DNA" NATURE, vol. 397, 18 February 1999, pages 579-583, XP002097746 LONDON GB | 1-3 |
| P,X | HENDRICH B ET AL: "Identification and characterization of a family of mammalian methyl-CpG binding proteins." MOLECULAR AND CELLULAR BIOLOGY, (1998 NOV) 18 (11) 6538-47. JOURNAL CODE: NGY. ISSN: 0270-7306., XP002097747 United States see the whole document --- -/- | |
| <input checked="" type="checkbox"/> Further documents are listed in the continuation of box C. <input checked="" type="checkbox"/> Patent family members are listed in annex. | | |
| * Special categories of cited documents : "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. "&" document member of the same patent family | | |
| Date of the actual completion of the international search | | Date of mailing of the international search report |
| 24 March 1999 | | 08/04/1999 |
| Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040. Tx. 31 651 epo nl, Fax: (+31-70) 340-3016 | | Authorized officer Espen, J |

INTERNATIONAL SEARCH REPORT

International Application No
PCT/CA 98/01059

| C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT | | |
|--|--|-----------------------|
| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
| X | SZYF M. ET AL.: "Mammalian cells contain a general (CpG) DNA demethylating activity" MOLECULAR BIOLOGY OF THE CELL, vol. 4, October 1993, page 74A XP002087767 * abstract 426 * | 1 |
| X | ----- SZYF M ET AL: "Ras induces a general DNA demethylation activity in mouse embryonal P19 cells." J BIOL CHEM, MAY 26 1995, 270 (21) P12690-6, XP002087765 UNITED STATES see the whole document | 1 |
| X | ----- SZYF M: "DNA methylation properties: consequences for pharmacology." TRENDS PHARMACOL SCI, JUL 1994, 15 (7) P233-8, XP002087766 ENGLAND see page 234, right-hand column; figures 1,2 | 1 |
| X | ----- WEISS A ET AL: "The role of DNA demethylation during development." GENES TO CELLS, (1997 AUG) 2 (8) 481-6. REF: 35 JOURNAL CODE: CUF. ISSN: 1356-9597., XP002097748 ENGLAND: United Kingdom see figures 3,4 | 1 |
| X | ----- WEISS A ET AL: "DNA demethylation in vitro: involvement of RNA" CELL, vol. 86, 1996, pages 709-718, XP002097749 NA US see the whole document | 1 |
| X | ----- WO 95 15373 A (UNIV MCGILL ;SZYF MOSHE (CA)) 8 June 1995 see claims 12,13 ----- | 1 |

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CA 98/01059

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

See continuation sheet
2. ☒ Claims Nos.: 19:22 (in part)
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
Said claims relate to antagonists, inhibitors or small molecules without giving a true technical characterization of the claimed matter. In consequence, the scope of said claims is ambiguous and, moreover, their subject-matter is vague and not sufficiently disclosed.
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Although claims 4-6,8,9,12,14,20-25,27,28 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition (applies where the subject-matter of said claims relates to an vivo treatment).
Although claim 29 is directed to a diagnostic method practised on the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

INTERNATIONAL SEARCH REPORT

Information on patent family members

national Application No

PCT/CA 98/01059

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---------------------|----------------------------|---------------------|
| WO 9515373 A | 08-06-1995 | CA 2110213 A | 31-05-1995 |
| | | AU 1061395 A | 19-06-1995 |
| | | CA 2177031 A | 08-06-1995 |
| | | EP 0734436 A | 02-10-1996 |
| | | EP 0889122 A | 07-01-1999 |
| | | JP 9506253 T | 24-06-1997 |
| ----- | | | |